sustainability

Special Issue Reprint

City 4.0

Urban Planning and Development in the Age of Digital Transformation

Edited by
Tan Yigitcanlar, Bo Xia, Tatiana Tucunduva Philippi Cortese
and Jamile Sabatini Marques

mdpi.com/journal/sustainability

MDPI

City 4.0: Urban Planning and Development in the Age of Digital Transformation

City 4.0: Urban Planning and Development in the Age of Digital Transformation

Editors

Tan Yigitcanlar
Bo Xia
Tatiana Tucunduva Philippi Cortese
Jamile Sabatini Marques

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Editors
Tan Yigitcanlar
Queensland University
of Technology
Brisbane
Australia

Bo Xia
Queensland University
of Technology
Brisbane
Australia

Tatiana Tucunduva Philippi Cortese
University Nove de Julho
Sao Paulo
Brazil

Jamile Sabatini Marques
University of São Paulo
Sao Paulo
Brazil

Editorial Office
MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Sustainability* (ISSN 2071-1050) (available at: https://www.mdpi.com/journal/sustainability/special_issues/City_4).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-7258-0094-0 (Hbk)
ISBN 978-3-7258-0093-3 (PDF)
doi.org/10.3390/books978-3-7258-0093-3

Contents

About the Editors

Tan Yigitcanlar

Tan Yigitcanlar is an eminent Australian researcher with international recognition and impact in the field of urban studies and planning. He is a Professor of Urban Studies and Planning at the School of Architecture and Built Environment, Queensland University of Technology, Brisbane, Australia. Along with this post, he holds the following positions: Honorary Professor at the School of Technology, Federal University of Santa Catarina, Florianopolis, Brazil; Director of the Australia-Brazil Smart City Research and Practice Network; Lead of QUT Smart City Research Group; and Co-Director of QUT City 4.0 Lab. He is a member of the Australian Research Council College of Experts.

Bo Xia

Bo Xia is a Professor of Construction Management in the School of Architecture and Built Environment, Queensland University of Technology, Brisbane, Australia. His research interests are in design–build, project delivery systems, construction management, sustainable buildings, and built environment, particularly for older people. His research received funding from nationally competitive grant schemes and findings of these projects have been published in internationally leading high impact journals and presented at mainstream international conferences. Bo Xia is a co-director of the City 4.0 Lab located at the School of Architecture and Built Environment, Queensland University of Technology.

Tatiana Tucunduva Philippi Cortese

Tatiana Tucunduva Phillippi Cortese holds a PhD and Master's in Science from the School of Public Health, University of São Paulo. She is a specialist in Environmental Law from USP Law School and Public Health School and graduated with a degree in Law from University Center UniFMU. Tatiana is a Professor at GVlaw, a Professor of the Academic Master's Program in Smart and Sustainable Cities–PPGCIS, and of the Professional Master's Program in Administration-Environmental Management and Sustainability–PPGeAS, at Uninove. She is an undergraduate Professor of Law at Uninove, a guest lecturer at Senac University Center, and author of the books *Smart and Sustainable Cities* and *Climate Changes: from Global to Local*. She works in the areas of Environmental and Urban Law; Public Policies and Environmental Management; Smart and Sustainable Cities; Sustainability Management; Climate Change; and interacting with interdisciplinary and multidisciplinary groups.

Jamile Sabatini Marques

Jamile Sabatini-Marques holds a PhD in Engineering and Knowledge Management from the Federal University of Santa Catarina, Brazil, where she spent the final year of her PhD studies at QUT under the supervision of Prof Tan Yigitcanlar. Following her PhD, she took up two consecutive postdoctoral research appointments at the University of São Paulo and Federal University of Santa Catarina. She is currently the Director of Innovation and Development of ABES (Brazilian Association of Software Companies), a member of the Brazilian Chamber of Information Technology, and a board member of the technology company Chipus. As an active member of the Australia-Brazil Smart City Research and Practice Network, she collaborates with QUT researchers.

Preface

The combination of the computer and distributed networks during recent decades has led to two 'Digital Revolutions' that today allow anyone to create, disseminate, and access any information anywhere, any time, and from any smart device. The birth of the 'Digital Age' is a result of a research ecosystem that was nurtured by government spending and military–industry–academia collaboration, along with the alliance of community organizers, communal-minded hippies, do-it-yourself hobbyists, and homebrew hackers. Strictly speaking, this collaborative creativity that defines the digital age included the collaboration between humans and machines—also known as 'collaborative intelligence'. In recent years, particularly the expansion of the Fourth Industrial Revolution (or Industry 4.0) to cities has disrupted the way in which cities are planned and developed as well as having generated a new city conceptualization—i.e., City 4.0. This new smart city blueprint, City 4.0, aims to leverage the power of engaged and connected citizens, digital technology, and data to ensure and enhance the quality of urban life, productivity, and sustainable development. In other words, City 4.0 is a city that utilizes technological developments and digitalization to transform local public services and the local economy to produce sustainable and desired urban, environmental, and societal outcomes for all. This reprint further elaborates the concept of City 4.0.

Tan Yigitcanlar, Bo Xia, Tatiana Tucunduva Philippi Cortese, and Jamile Sabatini Marques
Editors

sustainability

Editorial

City 4.0: Digital Transformation of Urban Settlements

Tan Yigitcanlar [1,*], Bo Xia [1], Tatiana Tucunduva Philippi Cortese [2] and Jamile Sabatini-Marques [3]

[1] City 4.0 Lab, School of Architecture and Built Environment, Faculty of Engineering, Queensland University of Technology, 2 George Street, Brisbane, QLD 4000, Australia; paul.xia@qut.edu.au
[2] Graduate Program in Smart and Sustainable Cities, University Nove de Julho, Rua Vergueiro, São Paulo 01525-000, Brazil; taticortese@gmail.com
[3] Institute of Advanced Studies, University of São Paulo, São Paulo 05508-060, Brazil; jamile.sabatini@abes.org.br
* Correspondence: tan.yigitcanlar@qut.edu.au; Tel.: +61-7-3138-2418

The combination of the computer and distributed networks during the recent decades has led to two 'Digital Revolutions' that today allow anyone to create, disseminate, and access any information anywhere, any time, and from any smart device. The birth of the 'Digital Age' is a result of a research ecosystem that was nurtured by government spending and military–industry–academia collaboration, along with the alliance of community organizers, communal-minded hippies, do-it-yourself hobbyists, and homebrew hackers. Strictly speaking, this collaborative creativity that defines the digital age included the collaboration between humans and machines—also known as 'collaborative intelligence'.

This interaction has changed the way in which some services are delivered. For instance, today, the world's largest taxi company, Uber, owns no vehicles; the world's most popular media owner, Facebook, creates no content; the world's most valuable retailer, Alibaba, has no inventory; and the world's largest accommodation provider, Airbnb, owns no real estate. In addition, owing to rapid developments in the digital age, technology is widely seen as an effective apparatus to help us solve some of the most challenging problems the world is currently facing, particularly when our impact is strongly considered alongside our technological capabilities.

At present, unexceptionally, all parts of the world are confronted with various environmental, social, health, and economic crises—e.g., life-threatening natural disasters, the loss of biodiversity, the destruction of natural ecosystems, regional disparities, socioeconomic inequity, pandemics, and digital and knowledge divides that are mainly caused by a rapid population increase and an expansion of resource consumption, combined with industrialization, urbanization, mobilization, agricultural intensification, and excessive consumption-driven lifestyles. The rapid advancement in digital technologies provides us with the hope that the impacts of global-scale environmental, social, and economic crises can be eased with the help of appropriate technology.

Furthermore, in recent years, the expansion of the Fourth Industrial Revolution (or Industry 4.0) to cities has disrupted the way in which cities are planned and developed as well as having generated a new city conceptualization—i.e., City 4.0. This new smart city blueprint, City 4.0, aims to leverage the power of engaged and connected citizens, digital technology, and data to ensure and enhance the quality of urban life, productivity, and sustainable development. In other words, City 4.0 is a city that utilizes technological developments and digitalization to transform local public services and the local economy to produce sustainable and desired urban, environmental, and societal outcomes for all.

Following this guest editorial backdrop, this Special Issue ('City 4.0: Urban Planning and Development in the Age of Digital Transformation') contains the following 10 papers that review, empirically explore, or theoretically expand the City 4.0 concept and practice from the various angles of urban planning and development (or, in simple terms, urban

Citation: Yigitcanlar, T.; Xia, B.; Cortese, T.T.P.; Sabatini-Marques, J. City 4.0: Digital Transformation of Urban Settlements. *Sustainability* **2024**, *16*, 671. https://doi.org/10.3390/su16020671

Received: 7 January 2024
Accepted: 11 January 2024
Published: 12 January 2024

planning and development in the age of Industry 4.0, which is also referred to as planning the digital transformation of urban settlements).

In Paper #1, entitled 'Understanding City 4.0: A Triple Bottom Line Approach', authors expand our understanding of City 4.0 by elaborating it from three diverse but interrelated perspectives—namely, societal, environmental, and economic lenses or domains (also known as the triple bottom line approach)—and highlight the key City 4.0 themes—namely, circularity, adaptability, livability, accessibility, authenticity, and responsibility. The study findings inform researchers, local and regional authorities, and urban planners about the rising importance of the notion of City 4.0 and its prospective research areas.

In Paper #2, entitled 'Synergy of Patent and Open-Source-Driven Sustainable Climate Governance under Green AI: A Case Study of TinyML', authors provide a conceptual expansion of climate and environmental policy in open synergy with proprietary and open source Tiny Machine Learning (TinyML) technology, and offer an industry collaborative and policy perspective on the issue, through using differential game models. This study finds that collaboration and sharing can lead to the implementation of green AI, reducing energy consumption and carbon emissions, and helping to fight climate change and protect the environment.

In Paper #3, entitled 'Does the Digital Economy Promote Coordinated Urban–Rural Development? Evidence from China', authors investigate whether and how the digital economy affects coordinated urban–rural development by using a panel data model, a Spatial Durbin Model (SDM), and a mediating effects model. The study findings provide a reference for China and other developing countries similar to China on how to promote coordinated urban and rural development in the development process of the digital economy.

In Paper #4, entitled 'Smarter Sustainable Tourism: Data-Driven Multi-Perspective Parameter Discovery for Autonomous Design and Operations', authors offer an approach for leveraging big data and deep learning to discover holistic, multi-perspective (e.g., local, cultural, national, and international), and objective information on a subject. This study develops a machine learning pipeline to extract parameters from the academic literature and public opinions on Twitter, providing a unique and comprehensive view of the industry from both academic and public perspectives.

In Paper #5, entitled 'Does Cross-Border E-Commerce Promote Economic Growth? Empirical Research on China's Pilot Zones', authors investigate whether the construction of China's Cross-Border E-Commerce (CBEC) comprehensive pilot zones can promote economic growth and social sustainable development. This study employs the difference-in-differences method to test the impact of the establishment of CBEC comprehensive pilot zones on economic growth and discusses the impact mechanism.

In Paper #6, entitled 'Augmenting Community Engagement in City 4.0: Considerations for Digital Agency in Urban Public Space', authors develop four augmented reality experiences to learn more about the potential for this technology to transform community engagement practices in the context of City 4.0. The study findings highlight the value of augmented reality's affordances to bring to light new interactions between community members and project stakeholders.

In Paper #7, entitled 'Transformation of Industry Ecosystems in Cities and Regions: A Generic Pathway for Smart and Green Transition', the author examines the pathways towards a digital and green transition by assessing a generic pathway for the transformation of industry ecosystems in cities and regions based on processes of prioritization, ecosystem identification, and platform-based digital and green transition. This study generates insights into pathways, priorities, and methods. This enables public authorities and business organizations to master the current industrial transformation of cities.

In Paper #8, entitled 'Automatically Generating Scenarios from a Text Corpus: A Case Study on Electric Vehicles', authors propose to further automate the process of scenario generation by guiding pre-trained deep neural networks (i.e., BERT) through simulated conversations to extract a model from a corpus. Their case study on electric vehicles shows that the approach yields similar results to previous work while almost eliminating the need for manual involvement in model building, thus focusing human expertise on the final stage of crafting compelling scenarios.

In Paper #9, entitled 'Musawah: A Data-Driven AI Approach and Tool to Co-Create Healthcare Services with a Case Study on Cancer Disease in Saudi Arabia', authors propose a data-driven artificial intelligence (AI)-based approach (called Musawah) to automatically discover healthcare services that can be developed or co-created by various stakeholders using social media analysis. This study emphasizes that open service and value healthcare systems based on freely available information revolutionize healthcare in manners similar to the open source revolution by using information made available by the public, the government, third and fourth sectors, or others, allowing new forms of preventions, cures, treatments, and support structures.

In Paper #10, entitled 'Crowdsourcing Public Engagement for Urban Planning in the Global South: Methods, Challenges and Suggestions for Future Research', authors provide a comprehensive overview of the crowdsourcing methods applied to public participation in urban planning in the Global South, as well as the technological, administrative, academic, socio-economic, and cultural challenges that could affect their successful adoption. This study puts forward important suggestions for both researchers and practitioners, where crowdsourcing has great benefits for the development of sustainable cities in the Global South.

This Special Issue, with its 10 paper contributions, generates new insights by investigating City 4.0 from various disciplinary angles to the digital transformation of urban settlements. We, the guest editors, believe that this Special Issue will be an important repository of relevant information, material, and knowledge to support research, policymaking, practice, and transferability of experiences regarding City 4.0.

Before we close, we wish to thank the authors of this Special Issue's papers for accepting our invitation and submitting and revising their manuscripts within a short time frame. Moreover, we would like to thank the referees for their thorough and timely reviews, as well as the journal's Assistant Editor, Ms. Sunnie Wei, for inviting us to serve as the Guest Editors of this Special Issue.

Author Contributions: Conceptualization and writing—original draft preparation, T.Y., B.X., T.T.P.C. and J.S.-M. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

sustainability

Communication

Understanding City 4.0: A Triple Bottom Line Approach

Tan Yigitcanlar [1,*], Bo Xia [1], Tatiana Tucunduva Philippi Cortese [2] and Jamile Sabatini-Marques [3]

[1] City 4.0 Lab, School of Architecture and Built Environment, Faculty of Engineering, Queensland University of Technology, 2 George Street, Brisbane, QLD 4000, Australia; paul.xia@qut.edu.au
[2] Graduate Program in Smart and Sustainable Cities, University Nove de Julho, Rua Vergueiro, São Paulo 01525-000, Brazil; tatianatpc@uni9.pro.br
[3] Institute of Advanced Studies, University of São Paulo, São Paulo 05508-060, Brazil; jamile.sabatini@abes.org.br
* Correspondence: tan.yigitcanlar@qut.edu.au; Tel.: +61-7-3138-2418

Abstract: Rapid urbanization and population increase, along with remarkable technological advances, have accelerated the speed of digital transformation, or at least the need for it, in our cities. Whilst being smart and sustainable is seen somewhat as an ideal city quality globally, a new city concept has emerged—the so-called City 4.0—that combines Industry 4.0 and Society 4.0 in the context of smart cities. While there is growing literature on the topic, there is limited understanding of City 4.0. This communication piece aims to bring clarity to City 4.0 by elaborating it from three diverse but interrelated perspectives—namely, societal, environmental, and economic lenses or domains (also know as the triple bottom line approach)—and highlights the key City 4.0 themes—namely, circularity, adaptability, livability, accessibility, authenticity, and responsibility. The methodological approach includes a thorough appraisal of the current City 4.0 literature. This communication paper informs researchers, local and regional authorities, and urban planners on the rising importance of the notion of City 4.0 and its prospective research areas.

Keywords: City 4.0; Industry 4.0; Society 4.0; smart city; sustainable development; knowledge-based development; urban innovation; responsible innovation; platform urbanism; digital transformation

Citation: Yigitcanlar, T.; Xia, B.; Cortese, T.T.P.; Sabatini-Marques, J. Understanding City 4.0: A Triple Bottom Line Approach. *Sustainability* **2024**, *16*, 326. https://doi.org/10.3390/su16010326

Academic Editor: Marc A. Rosen

Received: 15 November 2023
Revised: 25 December 2023
Accepted: 28 December 2023
Published: 29 December 2023

1. Introduction: Defining City 4.0

At present, way over half of the world's population resides in urban localities, making cities undoubtedly the epicenters of socioeconomic activities [1]. In some countries, the urbanization rate surpasses the 90% mark, such as Australia, the Netherlands, and the UK [2]. Along with cities being the primary generators of gross domestic product and innovation, rapid urban growth also brings challenges related to sustainability, quality of life, productivity, and urban management [3]. More specifically, we face the daunting task of ensuring that the urban centers of tomorrow do not exacerbate environmental degradation, social inequalities, and resource scarcity but rather serve as catalysts for progress toward a more sustainable, livable, and equitable world.

The major challenges we face today include, but not limited to the following: (a) The need to offer solutions to anthropogenic problems that cities have been causing; (b) The need to respond to the increasing socio-technological and infrastructural needs of growing urban regions; (c) The need to address the anthropogenic climate emergency; (d) The need to cater to the aging population's needs by using state-of-the art planning, design, construction, and management; (e) The need to offer cutting-edge and innovative solutions to improve the quality of life of residents; and (f) The need to prepare for digital transformation and technological disruption [4–6].

City 4.0 is the response to such challenges [7,8]. In a nutshell, City 4.0 is a city that utilizes technological developments and digitalization to transform local public services and the local economy to produce sustainable and desired urban, environmental, and societal outcomes for all [9]. It represents a vision where cities are smart, sustainable, inclusive, well-governed, and resilient [10,11]. As our planet undergoes an unprecedented wave of intense urbanization processes, the concept of City 4.0, following the smart city's popularity and foundations, has emerged as a beacon of hope on the horizon [12]; it is becoming a vision for many cities in the 21st century that integrate technology, innovation, and sustainability to address the myriad challenges of rapid urbanization [13,14].

City 4.0 envisions urban environments where technology enhances resource efficiency, connectivity improves quality of life, and data-driven decision-making transforms governance. It represents the convergence of urban development and the digital age, offering innovative solutions to age-old urban problems [5,15,16]. By integrating various technological advancements like Internet-of-Things (IoT), artificial intelligence (AI), big data, and sustainable practices to manage resources, transportation, energy, and services within a city, City 4.0 aims to create smarter, more sustainable, more efficient, and more livable and prosperous cities [6,17].

In terms of the differences between smart city and City 4.0, both refer to urban development concepts, but each of them emphasizes different aspects and stages of technological advancement within a city. In essence, while smart city focuses on using technology to improve specific aspects of city life, City 4.0 represents a more advanced stage, encompassing a broader and more deeply integrated innovation and technological ecosystem aimed at creating a highly connected, intelligent, and efficient urban environment. City 4.0 is often seen as the next phase of smart city development, incorporating more advanced and disruptive technologies to transform urban living and governance [5,18].

As a result, the concept of City 4.0 represents a paradigm shift in urban planning and development, "leveraging the power of engaged and connected citizens, digital technology, and data to ensure and enhance the quality of urban life, productivity, and sustainable development" [9,19]. Although such a paradigm shift is occurring at the moment, there is a knowledge gap regarding what City 4.0 is and what its key characteristics are. Tackling this issue is the rationale behind this communication piece.

This is a communication paper, as opposed to a research article, we aim to contribute to the efforts in bridging this knowledge gap by sharing our views on the topic thorough an appraisal of the current City 4.0 literature and our extensive experience and expertise in the field of smart and sustainable urbanism. In this communication piece, we explore City 4.0 from the perspectives of societal, environmental, and economic lenses (a triple bottom line approach), and also introduce the six prominent themes—circularity, adaptability, livability, accessibility, authenticity, and responsibility—of City 4.0.

2. City 4.0 from the Societal Lens

The societal lens of City 4.0 delves into this new city type's societal benefits and challenges, and the transformative potential it holds for urban societies, including the rapidly growing aging population. City 4.0 brings profound societal benefits, influencing various aspects of urban life through the adoption of various digital technologies and data-driven decision-making [20].

First, City 4.0 enables active citizen engagement, since digital platforms, mobile applications, and participatory mechanisms enable citizens to provide feedback, report issues, and actively participate in local initiatives to shape and better control their urban environment [21]. This active involvement fosters a sense of ownership, belongingness, accountability, and community cohesion among citizens, maximizing their contribution to community development and generating more sustainable solutions [9,22].

Second, City 4.0 promotes inclusive governance and community, where decision-making is transparent and responsive to the diverse needs of stakeholders, particularly for community citizens. This inclusivity ensures that marginalized and underrepresented groups, especially those vulnerable and disadvantaged populations in communities, have a voice in urban development, leading to more equitable outcomes [23].

Third, City 4.0 enhances the delivery of public services and infrastructure for the improvement in the quality of urban life. With data-driven insights and support from AI, governments can optimize services and infrastructure provision such as transportation, healthcare, education, and public safety, leading to a smart society that benefits all residents with improved efficiency and effectiveness [24]. To improve these deliveries, it is necessary to have good data to work on.

While City 4.0 holds immense promise for our current and future society, it is not without its challenges and considerations, including, but not limited, to the following:

Privacy and Cybersecurity: The extensive data collection and analysis required for City 4.0 raises concerns about privacy and cybersecurity. Safeguarding citizen data and ensuring data privacy regulations are adhered to is crucial to maintain public trust. Additionally, as cities become increasingly interconnected, they become more vulnerable to cyber threats. Robust cybersecurity measures are necessary to protect critical infrastructure and citizen data [25]. Robust rules and regulations, such as the General Data Protection Law, must be implemented to ensure that citizens' rights are protected [26].

Digital Divide: While City 4.0 envisions a connected and engaged citizenry, it also highlights the digital divide that exists within our societies because not all citizens have equal access to digital technology or possess the necessary digital literacy. Addressing this divide is crucial to ensure that City 4.0 benefits all segments of the population [27]. This is to say, technology must be accessible and beneficial to all citizens, regardless of their socioeconomic status, to avoid deepening social inequalities.

Public Acceptance: The changes brought by City 4.0 may not find universal acceptance among all residents. There may be resistance to new technologies, concerns about job displacement, or resistance to the disruption of established routines. In addition, urban citizens from different cultures and societies may have varying levels of acceptance and readiness for digital transformation [28].

Keeping these issues in mind, City 4.0 initiatives should be sensitive to all these factors to ensure successful implementation. It is worth mentioning that City 4.0 brings great opportunities and innovative solutions for addressing issues related to the rapidly growing aging population. The aging population is a demographic trend that is reshaping societies worldwide. By 2050, it is estimated that over 22% of the global population will be aged 60 or older [29]. This demographic shift poses unique challenges for urban planners and policymakers, but City 4.0 offers promising solutions.

For example, the emphasis of City 4.0 on data-driven decision-making can lead to the creation of age-friendly infrastructures, which include well-designed public spaces, accessible transportation options, and smart homes that cater to the specific needs of older citizens. In addition, digital technology in City 4.0 facilitates telemedicine and remote healthcare services, making it easier for older adults to access medical care without the need for extensive travel [30].

Additionally, given that loneliness and social isolation are significant concerns for the aging population [31], City 4.0 platforms can foster social connections by facilitating virtual meetups, providing information about local social activities, and encouraging intergenerational interactions. Finally, considering that older adults often face challenges with transportation and mobility, especially in sprawling urban areas, the smart transportation systems of City 4.0 can offer on-demand, accessible, and personalized transportation options, ensuring that seniors can remain mobile and independent [32].

To summarize, the emergence of City 4.0 carries a range of important societal implications, all of which are significant and multifaceted, influencing various aspects of urban life and society. On the one hand, it offers opportunities for enhanced citizen engagement, improved governance, sustainability, and better quality of life. On the other hand, it also presents challenges related to concerns regarding privacy, cybersecurity, the digital divide, and public acceptance. Meanwhile, it is worth noting that City 4.0 has profound implications for an aging population, the majority of which will age in cities. By building age-friendly infrastructures, improving healthcare access, promoting social inclusion, and enhancing transportation options, City 4.0 can significantly improve the quality of life for older citizens, although it is also crucial to address those similar issues like the digital divide, privacy concerns, and digital literacy to ensure at City 4.0 truly benefits all members of society, regardless of age.

3. City 4.0 from the Environmental Lens

City 4.0′s environmental perspective has a strong sustainability and sustainable urban development focus. As outlined in the UN's Sustainable Development Goals (SDGs), sustainability encompasses environmental, social, and economic well-being [33,34]. Within this framework, we examine how City 4.0 contributes to achieving these global objectives.

Technology and Innovation: City 4.0 adopts advanced technologies such as AI, IoT, and big data analytics to create more efficient, sustainable, and resilient urban environments. In addition, the adoption of these technologies can optimize transportation, energy consumption, and waste management, contributing to sustainable resource use and reduced emissions in cities [35].

Smart Infrastructure and Services: An essential aspect of City 4.0 is the development of smart infrastructure and services, such as intelligent transportation systems, renewable energy generation, waste reduction strategies, and other urban services [36]. These innovations not only enhance the livability of cities (as discussed in the Society Lens of City 4.0) but also significantly reduce their environmental footprint [37].

Data-Driven Decision-Making: In the context of City 4.0, the emphasis is on the environmentally responsible collection and analysis of extensive data to enhance the sustainability of urban services and the efficient allocation of resources. This highlights how data-driven approaches in urban planning can lead to more eco-conscious governance, offering insights into the environmental challenges and opportunities associated with this approach [38]. In City 4.0, data will promote innovation and technological ecosystems and governments can include the smart mind society to contribute to solve city problems and to have more transparency in the process [39].

Environmental Benefits: City 4.0 can mitigate environmental degradation, given that sustainable urban planning can enhance green spaces, preserve biodiversity, improve air quality, and promote climate resilience. The objective is to understand how City 4.0 contributes to sustainability within city boundaries and on a global scale [40,41].

Social Equity and Inclusion: Sustainable development encompasses social dimensions, and our special edition recognizes their importance in the context of City 4.0. We delve into the role City 4.0 plays in providing affordable housing, bridging the digital divide, and ensuring community engagement. The vision of inclusive, equitable, and diverse cities is explored in detail, emphasizing the significance of balancing technological advancements with social progress [42,43].

Therefore, City 4.0 supports sustainable development because cities can reduce their environmental footprint, lower energy consumption, minimize waste and mitigate the impact of climate change by harnessing data and digital technology [44]. Regulating the carbon emission trading policies and promoting innovation in green technologies is crucial for City 4.0 in addressing the climate emergency [45]. This not only benefits the present generations, but also contributes to a more sustainable future for urban societies.

Effective policies and regulations are pivotal in guiding sustainable urban development and governments at all levels must play a central role in incentivizing and regulating the transition to City 4.0. Public awareness and community engagement will be equally vital, as informed and engaged citizens are instrumental in driving the transition to more sustainable, equitable, and resilient cities.

In summary, viewed through the lens of sustainable development, the concept of City 4.0 offers both a vision and a roadmap toward a brighter and more equitable urban future. With this communication piece, we hope to stimulate conversations, inspire action, and provide a platform for knowledge sharing among researchers, policymakers, and practitioners as we collectively embark on the journey to City 4.0, a vision where our cities will not only flourish, but also be the cornerstones of a more sustainable and inclusive global society.

4. City 4.0 from the Economic Lens

The economic lens of City 4.0 adopts urban innovation principles where, in this regard, knowledge-based collective platform urbanism paves the way for a more economically sustainable, resilient, and inclusive urban future [39]. Knowledge-based collective platform urbanism, a key component of City 4.0, proposes a revolution in the way cities manage their development [46]. At the heart of this strategy is the establishment of an innovative environment that values cooperation between diverse actors—from government entities and the business sector to academia and citizens. This confluence of visions and expertise makes it possible to create solutions to complex contemporary urban problems.

The concept of City 4.0 expands the traditional idea of urbanism by emphasizing the importance of technology and the use of data in improving urban infrastructures. It focuses on capturing and analyzing real-time data to improve urban systems such as transportation and energy management, as well as human resource development, smart marketing, and smart technology adoption [47]. This is complemented by knowledge-based collective platform urbanism, which highlights the role of collective knowledge in solving complex urban challenges. Both approaches agree on the centrality of data, but while City 4.0 focuses on optimization and efficiency, knowledge-based collective platform urbanism sees data as a public resource that must be democratized to foster innovation [48].

The importance of entrepreneurship and innovation in knowledge-based economic development is undeniable. To promote this integration and economic development, cities must have universities that train a specialized and entrepreneurial workforce, as well as support infrastructures such as incubators, accelerators, and technology parks [49]. Legislation and public policies must also favor the growth of these emerging companies, whether through investment funds or tax incentives. Furthermore, political ties, environmental regulations, and corporate innovation should be considered in combination [50].

In the context of City 4.0, creating a robust innovation ecosystem requires a holistic approach that integrates the UN's three pillars of sustainability: "economic development, social development, and environmental protection" [51]. This ecosystem must operate in four main domains: socio-cultural, spatial, institutional, and economic. Each domain has its pillars; however, when we look at the economic-domain-emphasizing aspects, such as knowledge, creativity, fostering, innovation, and competitiveness, we see cities being able to generate economic development that is both sustainable and knowledge-based [52,53].

A crucial pillar of this vision is the emphasis on open data. By making this data accessible to a wide range of stakeholders, a stimulus is created for the development of new applications and services aimed at improving the quality of urban life. These innovations can cover areas as diverse as transportation, energy management, public safety, and health [54].

Data management plays a central role in this ecosystem, providing the necessary stimulus to drive innovation [55]. These data, seen as commons or common goods, have immense power in society and can be used to reorganize cities in more efficient and just ways. These commons also include other natural or cultural resources and are fundamental to sustainable development.

Promotion, whether through favorable tax policies, subsidies, or support programs for entrepreneurs, plays a crucial role in the development or transformation of City 4,0. Cities can foster an environment where startups flourish and innovation is valued by encouraging them to generate innovations. This approach does not just benefit companies; it creates a dynamic where all citizens benefit from technological advances and innovative solutions. Especially when fostered by open and big data, the intelligence of the collective is used to contribute and bring invaluable solutions to cities.

Therefore, the union of City 4.0 with knowledge-based collective platform urbanism, and the importance of entrepreneurship and innovation, offer a transformative paradigm for the future of urban development. These are not just theoretical concepts, but a call to action that demands a radical reimagining of how cities function and meet the needs of their inhabitants. The possibilities are endless and, even in the face of significant challenges, these approaches point towards more resilient, inclusive, and innovative cities.

5. City 4.0's Prominent Themes

When elaborating on City 4.0, it is also important to present the prominent themes this new city conceptualization comprises. Thus, this section concentrates on this issue. In City 4.0, which aims to utilize technological developments and digitalization to transform local public services and local economies to produce sustainable and desired urban, environmental, and societal outcomes for all, the following six themes become prominent—circularity, adaptability, livability, accessibility, authenticity, and responsibility [56]. These interrelated key themes are briefly introduced below and a framework with some of the exemplar approaches to these key themes are graphically presented with some examples in Figure 1.

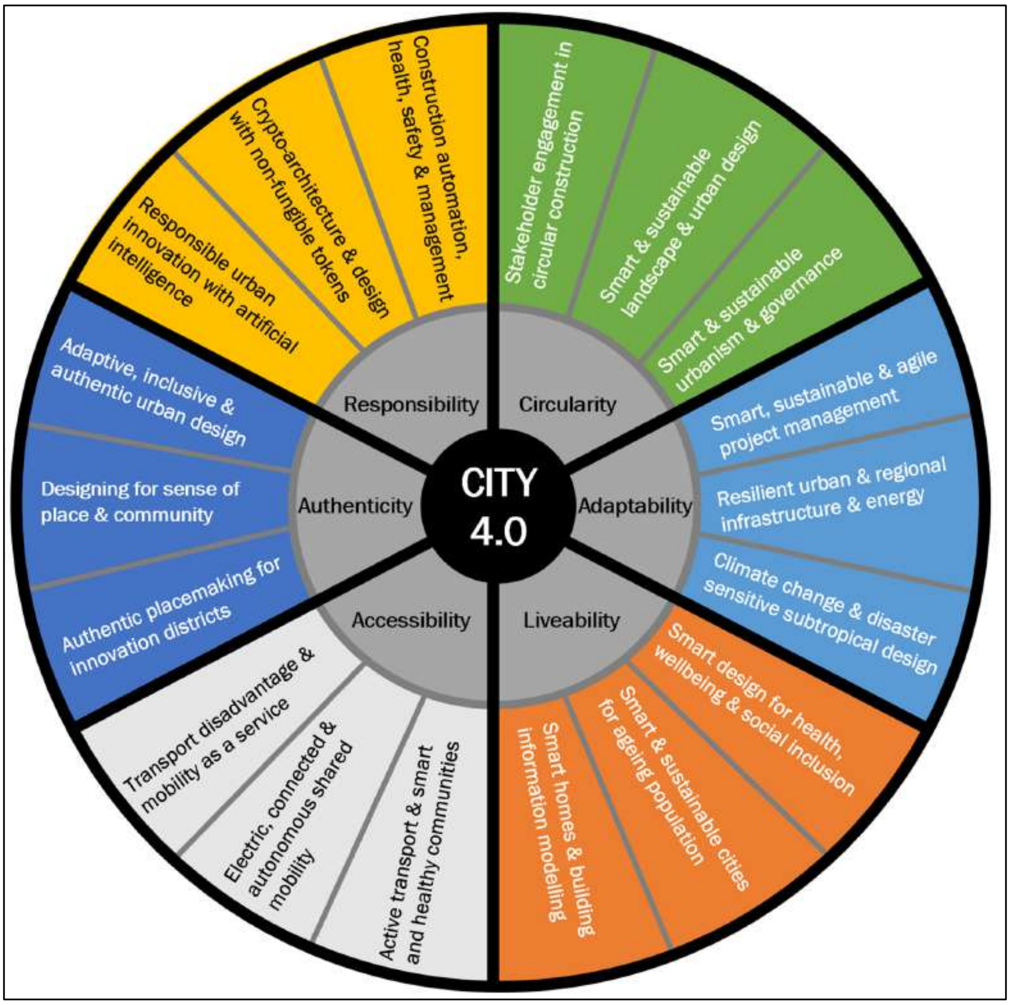


Figure 1. Framework of City 4.0's prominent themes.

In the core ring of Figure 1 sits City 4.0; the middle-ring contains prominent themes of City 4.0, while the outer-ring includes some project examples for a clearer understanding of each theme. The six themes are concisely explained below.

Circularity: This theme focuses on economic, social, technological, and environmental systems that aim to eliminate waste, maximize the reusing of resources, and contribute to the sustainable development efforts of City 4.0.

Adaptability: This theme focuses on boosting the ability to make necessary adjustments in ecological, social, technological, and economic systems in response to actual or expected climatic stimuli and their effects or impacts on City 4.0.

Livability: This theme focuses on enhancing the conditions of a decent life for all inhabitants of cities, regions, and communities, including their physical and mental wellbeing, in City 4.0.

Accessibility: This theme focuses on informing the practice of making information, activities, opportunities, and environments sensible, meaningful, and equitably usable for as many people as possible in City 4.0.

Authenticity: This theme focuses on forming and maintaining spaces, places, and communities that are genuine and distinctive and contain recognizable social–cultural and natural characteristics and identities in City 4.0.

Responsibility: This theme focuses on informing governance decisions, including in technology adoption and utilization, through ethical, accountable, trustworthy, explainable, and equitable frameworks in City 4.0 [56].

The presence of each of these six prominent themes as successfully applied projects in a city (see the outer ring of Figure 1 for project examples), at scale, will initiate the transformation journey of that locality into a smarter and more sustainable city and will thus pave the way to the emergence of City 4.0.

6. Concluding Remarks: City 4.0's Prominent Research Areas

This communication piece aims to bring clarity to City 4.0, which is seen as the next phase of smart city development, by elaborating on it from the three diverse but interrelated perspectives—namely, the societal, sustainability, and economic lenses or domains (also known as the triple bottom line approach). The piece also highlighted the prominent themes in City 4.0—circularity, adaptability, livability, accessibility, authenticity, and responsibility.

Despite our communication piece shedding some light on the notion of City 4.0, the notion is still in its infancy. We hence call for in-depth research, effective implementation, and thoughtful consideration of the societal, environmental, and economic implications to realize the full potential of City 4.0, while tackling its associated challenges.

Additionally, we advocate and encourage collaboration among academics, governments, communities, business, and industry professionals to ensure that our future cities are smart, inclusive, sustainable, and responsive to the diverse needs of urban populations. A concentrated effort is needed to reach the potential of City 4.0 in shaping the future of our urban centers and to transform the vision of City 4.0 into a tangible reality. This also means fostering global collaboration to address the challenges of urbanization and climate change. As cities continue to evolve and embrace the possibilities of the digital age, City 4.0 represents a promising vision for the future of urban life, one that is smart, sustainable, inclusive, well-governed, and resilient.

Global collaboration for urbanization and climate change is imperative to justify the vision of City 4.0. As urbanization accelerates worldwide, and the impacts of climate change become more pronounced, cities must unite to face common challenges. By fostering global collaboration, cities can share best practices and innovative solutions that transcend geographical boundaries. This collaborative approach enables the exchange of knowledge, experiences, and technological advancements, paving the way for the development of smart, sustainable, and resilient cities. In a world interconnected by shared environmental concerns, collaboration is the most important management tool for City 4.0, allowing cities to collectively navigate the complexities of urban growth and climate resilience.

Furthermore, City 4.0 needs to use collective intelligence to solve existing problems, bringing citizens closer. From this perspective, collaborative intelligence—where humans and AI join forces—is critical [57]. Moreover, data are the new oil and, hence, through open data, City 4.0 can promote entrepreneurship through hackathons, contributing to digital transformation and responsible innovation [58]. Furthermore, cities have the challenge of promoting education and inclusion for this new world—facing many societal, economic, governance and environmental problems— and, to do so, they need to provide digital network infrastructure and cybersecurity [59,60].

In this communication, we also propose public policy recommendations for achieving City 4.0. From the societal perspective, policies should be centered on safeguarding citizen data and ensuring data privacy, giving different groups of citizens equal access to digital technology and improving public acceptance of City 4.0, especially regarding readiness for digital transformation. From the perspective of the economic lens for City 4.0, it is important that there are regulations for opening up data through public policies, and that these data are clean and clear. As a public policy factor, guaranteeing a quality and secure internet connection is fundamental for achieving City 4.0 status. In a nutshell, urban policies should concentrate on initiating the transformation towards human-centered cities, focusing on social, economic, and environmental aspects, seeking transparency in management and utilizing technology and collective intelligence for city development.

To achieve the ideal state of City 4.0, we hence call for in-depth research, effective implementation, and thoughtful consideration of the societal, environmental, and economic implications in order to realize the full potential of City 4.0 while tackling its associated challenges. In addition, we advocate for and encourage global collaboration among academics, governments, communities, business, and industry professionals to ensure that our future cities are smart, inclusive, sustainable, and responsive to the diverse needs of urban populations.

We also strongly advocate that, with all above-mentioned factors—along with a sound and balanced triple bottom line approach (societal, environmental, and economic) with a focus on the circularity, adaptability, livability, accessibility, authenticity, and responsibility aspects of our cities—City 4.0 (utilizing digitalization as a disruptive force in cities) has the capacity to generate development based on knowledge and sustainability, whilst offering smarter and more sustainable, livable, and prosperous futures. We also believe that the shared ideas in this communication paper will inform researchers, authorities, and urban planners on the raising importance of the City 4.0 notion, and will encourage the uptake of projects concentrating on increasing the circularity, adaptability, livability, accessibility, authenticity, and responsibility of our cities.

Lastly, we conclude this communication piece by advocating for the importance of prospective research on City 4.0, and mapping some of the relevant and important research areas—as shown in Figure 2—that would be of interest to researchers dedicated to investigating sustainable, healthy, and digital transformations and their disruptions in and implications for our cities, communities, industries, and the environment.

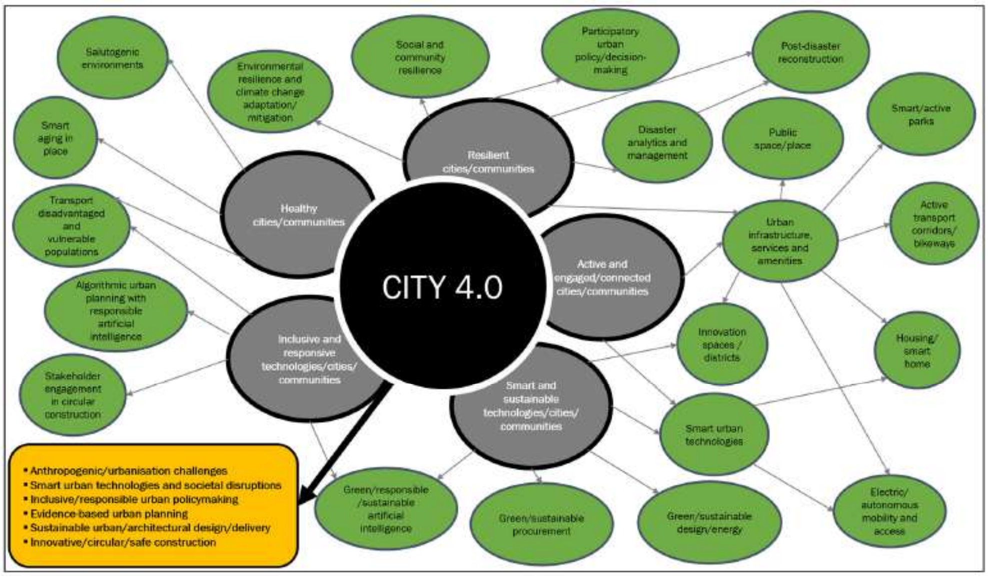


Figure 2. City 4.0’s prominent research areas.

Author Contributions: Conceptualization, data curation, analysis, writing—original draft preparation, and writing—editing and revision, T.Y., B.X., T.T.P.C. and J.S.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors thank the managing editor and the five anonymous referees for their invaluable comments on an earlier version of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Castells-Quintana, D.; Wenban-Smith, H. Population dynamics, rbanization without growth, and the rise of megacities. *J. Dev. Stud.* **2020**, *56*, 1663–1682. [CrossRef]
2. Dzator, J.; Acheampong, A.; Dzator, M. Urbanisation and sustainable development: Econometric evidence from Australia. In *Community Empowerment, Sustainable Cities, and Transformative Economies*; Chaiechi, T., Wood, J., Eds.; Springer: Singapore, 2022.
3. Zhang, X. The trends, promises and challenges of urbanisation in the world. *Habitat Int.* **2016**, *54*, 2241–2252. [CrossRef]
4. Chan, F.; Chan, H. Recent research and challenges in sustainable urbanisation. *Resour. Conserv. Recycl.* **2022**, *184*, 106346. [CrossRef]
5. Fernandez-Anez, V.; Fernández-Güell, J.; Giffinger, R. Smart city implementation and discourses: An integrated conceptual model: The case of Vienna. *Cities* **2018**, *78*, 4–16. [CrossRef]
6. Desdemoustier, J.; Crutzen, N.; Giffinger, R. Municipalities' understanding of the smart city concept: An exploratory analysis in Belgium. *Technol. Forecast. Soc. Chang.* **2019**, *142*, 129–141. [CrossRef]
7. Monaco, L. Looking for city 4.0: Two work in progress experiences in Zaragoza. In Proceedings of the Fab14+ Conference, Toulouse, France, 2–8 July 2018.
8. IIoT World. City Digitalization: Welcome to the city 4.0. Available online: https://www.iiot-world.com/smart-cities-buildings-infrastructure/smart-cities/digitalization-welcome-to-the-city-4-0 (accessed on 12 December 2023).
9. Hunter, M.; Soro, A.; Brown, R.; Harman, J.; Yigitcanlar, T. Augmenting community engagement in city 4.0: Considerations for digital agency in urban public space. *Sustainability* **2022**, *14*, 9803. [CrossRef]
10. Rosemann, M.; Becker, J.; Chasin, F. City 5.0. *Bus. Inf. Syst. Eng.* **2021**, *63*, 71–77. [CrossRef]
11. Becker, J.; Chasin, F.; Rosemann, M.; Beverungen, D.; Priefer, J.; Brocke, J.; Matzner, M.; del Rio Ortega, A.; Resinas, M.; Santoro, F.; et al. City 5.0: Citizen involvement in the design of future cities. *Electron. Mark.* **2023**, *33*, 10. [CrossRef]
12. Scribano, A. City and emotions. In *Cities, Capitalism and the Politics of Sensibilities*; Scribano, A., Camarena Luhrs, M., Cervio, A., Eds.; Palgrave Macmillan: Cham, Switzerland, 2021.
13. Yun, Y.; Lee, M. Smart city 4.0 from the perspective of open innovation. *J. Open Innov. Technol. Mark. Complex.* **2019**, *5*, 92. [CrossRef]
14. Boumale, B.; Tamine, R. Synergy between smart cities and industry 4.0 in public spaces: Bibliometric analysis. *Procedia Comput. Sci.* **2022**, *204*, 775–783. [CrossRef]
15. Lin, Y.; Hu, W.; Chen, X.; Li, S.; Wang, F. City 5.0: Towards spatial symbiotic intelligence via DAOs and parallel systems. *IEEE Trans. Intell. Veh.* **2023**, *8*, 3767–3770. [CrossRef]
16. Wolniak, R.; Gajdzik, B.; Grebski, W. The implementation of Industry 4.0 concept in smart city. *Sci. Pap. Silesian Univ. Technol. Organ. Manag. Ser.* **2023**, *178*, 753–770. [CrossRef]
17. Yigitcanlar, T.; Guaralda, M.; Taboada, M.; Pancholi, S. Place making for knowledge generation and innovation: Planning and branding Brisbane's knowledge community precincts. *J. Urban Technol.* **2016**, *23*, 115–146. [CrossRef]
18. Zubizarreta, I.; Seravalli, A.; Arrizabalaga, S. Smart city concept: What it is and what it should be. *J. Urban Plan. Dev.* **2016**, *142*, 04015005. [CrossRef]
19. D'Amico, G.; L'Abbate, P.; Liao, W.; Yigitcanlar, T.; Ioppolo, G. Understanding sensor cities: Insights from technology giant company driven smart urbanism practices. *Sensors* **2020**, *20*, 4391. [CrossRef] [PubMed]
20. Baltac, V. Smart cities: A view of societal aspects. *Smart Cities* **2019**, *2*, 538–548. [CrossRef]
21. Park, Y. Strategy for building smart city as a platform of the 4th industrial revolution. *J. Digit. Converg.* **2019**, *17*, 169–177.
22. Kinelski, G. Smart city 4.0 as a set of social synergies. *Pol. J. Manag. Stud.* **2022**, *26*, 92–106. [CrossRef]
23. Tan, S.; Taeihagh, A. Smart city governance in developing countries: A systematic literature review. *Sustainability* **2020**, *12*, 899. [CrossRef]

24. Suling, C.; Nurmandi, A.; Muallidin, I.; Purnomo, E.; Kurniawan, D. The use of AI to develop smart infrastructure in Indonesia. In Proceedings of the International Conference on Human-Computer Interaction, Aragón, Spain, 7–9 September 2022; Springer: Cham, Switzerland, 2022; pp. 208–217.
25. Elmaghraby, A.; Losavio, M. Cyber security challenges in smart cities: Safety, security and privacy. *J. Adv. Res.* **2014**, *5*, 491–497. [CrossRef]
26. Edwards, L. Privacy, security and data protection in smart cities: A critical EU law perspective. *Eur. Data Prot. Law Rev.* **2016**, *2*, 28–58. [CrossRef]
27. Kolotouchkina, O.; Barroso, C.; Sánchez, J. Smart cities, the digital divide, and people with disabilities. *Cities* **2022**, *123*, 103613. [CrossRef]
28. Habib, A.; Alsmadi, D.; Prybutok, V. Factors that determine residents' acceptance of smart city technologies. *Behav. Inf. Technol.* **2020**, *39*, 610–623. [CrossRef]
29. World Health Organization. Ageing. Available online: https://www.who.int/news-room/facts-in-pictures/detail/ageing#:~:text=The%20number%20of%20people%20aged,quickly%20than%20in%20the%20past (accessed on 16 October 2023).
30. Torku, A.; Chan, A.; Yung, E. Implementation of age-friendly initiatives in smart cities: Probing the barriers through a systematic review. *Built Environ. Proj. Asset Manag.* **2021**, *11*, 412–426. [CrossRef]
31. Peek, S.; Wouters, E.; Van Hoof, J.; Luijkx, K.; Boeije, H.; Vrijhoef, H. Factors influencing acceptance of technology for aging in place: A systematic review. *Int. J. Med. Inform.* **2014**, *83*, 235–248. [CrossRef] [PubMed]
32. Wang, C.; Steinfeld, E.; Maisel, J.; Kang, B. Is your smart city inclusive? Evaluating proposals from the US Department of Transportation's smart city challenge. *Sustain. Cities Soc.* **2021**, *74*, 103148. [CrossRef]
33. Kamruzzaman, M.; Hine, J.; Yigitcanlar, T. Investigating the link between carbon dioxide emissions and transport-related social exclusion in rural Northern Ireland. *Int. J. Environ. Sci. Technol.* **2015**, *12*, 3463–3478. [CrossRef]
34. Grossi, G.; Trunova, O. Are UN SDGs useful for capturing multiple values of smart city? *Cities* **2021**, *114*, 103193. [CrossRef]
35. Singh, S.; Sharma, P.; Yoon, B.; Shojafar, M.; Cho, G.; Ra, I. Convergence of blockchain and artificial intelligence in IoT network for the sustainable smart city. *Sustain. Cities Soc.* **2020**, *63*, 102364. [CrossRef]
36. Svítek, M.; Skobelev, P.; Kozhevnikov, S. Smart city 5.0 as an urban ecosystem of Smart services. In *Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future, Proceedings of SOHOMA, Shenzhen, China, 15–17 November 2019*; Springer: Singapore, 2019; pp. 426–438.
37. Selim, A.; Yousef, P.; Hagag, M. Smart infrastructure by (PPPs) within the concept of smart cities to achieve sustainable development. *Int. J. Crit. Infrastruct* **2018**, *14*, 182–198. [CrossRef]
38. Shahat-Osman, A.; Elragal, A. Smart cities and big data analytics: A data-driven decision-making use case. *Smart Cities* **2021**, *4*, 286–313. [CrossRef]
39. Almalki, F.; Alsamhi, S.; Sahal, R.; Hassan, J.; Hawbani, A.; Rajput, N.; Saif, A.; Morgan, J.; Breslin, J. Green IoT for eco-friendly and sustainable smart cities: Future directions and opportunities. *Mob. Netw. Appl.* **2023**, *28*, 178–202. [CrossRef]
40. Angelidou, M.; Psaltoglou, A.; Komninos, N.; Kakderi, C.; Tsarchopoulos, P.; Panori, A. Enhancing sustainable urban development through smart city applications. *J. Sci. Technol. Policy Manag.* **2018**, *9*, 146–169. [CrossRef]
41. Makieła, Z.; Stuss, M.; Mucha-Kuś, K.; Kinelski, G.; Budziński, M.; Michałek, J. Smart city 4.0: Sustainable urban development in the Metropolis GZM. *Sustainability* **2022**, *14*, 3516. [CrossRef]
42. Esmaeilpoorarabi, N.; Yigitcanlar, T.; Guaralda, M.; Kamruzzaman, M. Evaluating place quality in innovation districts: A Delphic hierarchy process approach. *Land Use Policy* **2018**, *76*, 471–486. [CrossRef]
43. Chen, T.; Ramon Gil-Garcia, J.; Gasco-Hernandez, M. Understanding social sustainability for smart cities: The importance of inclusion, equity, and citizen participation as both inputs and long-term outcomes. *J. Smart Cities Soc.* **2022**, *1*, 135–148. [CrossRef]
44. Blasi, S.; Ganzaroli, A.; De Noni, I. Smartening sustainable development in cities: Strengthening the theoretical linkage between smart cities and SDGs. *Sustain. Cities Soc.* **2022**, *80*, 103793. [CrossRef]
45. Gao, D.; Tan, L.; Mo, X.; Xiong, R. Blue sky defense for carbon emission trading policies: A perspective on the spatial spillover effects of total factor carbon efficiency. *Systems* **2023**, *11*, 382. [CrossRef]
46. Caprotti, F.; Chang, I.; Joss, S. Beyond the smart city: A typology of platform urbanism. *Urban Transform.* **2022**, *4*, 4. [CrossRef]
47. Gamayanto, I.; Nurhindarto, A. developing smart city 5.0 framework to produce competency. *J. Appl. Intell. Syst.* **2020**, *5*, 23–31. [CrossRef]
48. Van der Graaf, S.; Ballon, P. Navigating platform urbanism. *Technol. Forecast. Soc. Chang.* **2019**, *142*, 364–372. [CrossRef]
49. Ivaldi, E.; Penco, L.; Isola, G.; Musso, E. Smart sustainable cities and the urban knowledge-based economy: A NUTS3 level analysis. *Soc. Indic. Res.* **2020**, *150*, 45–72. [CrossRef]
50. Gao, D.; Li, Y.; Tan, L. Can environmental regulation break the political resource curse: Evidence from heavy polluting private listed companies in China. *J. Environ. Plan. Manag.* **2023**, 1–27. [CrossRef]
51. Fuso Nerini, F.; Sovacool, B.; Hughes, N.; Cozzi, L.; Cosgrave, E.; Howells, M.; Tavoni, M.; Tomei, J.; Zerriffi, H.; Milligan, B. Connecting climate action with other sustainable development goals. *Nat. Sustain.* **2019**, *2*, 674–680. [CrossRef]
52. Yigitcanlar, T.; Dur, F. Making space and place for knowledge communities: Lessons for Australian practice. *Australas. J. Reg. Stud.* **2013**, *19*, 36–63.
53. Appio, F.; Lima, M.; Paroutis, S. Understanding smart cities: Innovation ecosystems, technological advancements, and societal challenges. *Technol. Forecast. Soc. Chang.* **2019**, *142*, 1–14. [CrossRef]

54. Caragliu, A.; Del Bo, C. Smart innovative cities: The impact of smart city policies on urban innovation. *Technol. Forecast. Soc. Chang.* **2019**, *142*, 373–383. [CrossRef]
55. Ahmad, K.; Maabreh, M.; Ghaly, M.; Khan, K.; Qadir, J.; Al-Fuqaha, A. Developing future human-centered smart cities: Critical analysis of smart city security, data management, and ethical challenges. *Comput. Sci. Rev.* **2022**, *43*, 100452. [CrossRef]
56. City 4.0 Lab. Our Vision. Available online: https://research.qut.edu.au/citylab (accessed on 15 November 2023).
57. Wilson, J.; Daugherty, P. Collaborative Intelligence: Humans and AI Are Joining Forces. Available online: https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces (accessed on 15 November 2023).
58. Batty, M. Big data, smart cities and city planning. *Dialogues Hum. Geogr.* **2013**, *3*, 274–279. [CrossRef]
59. Barns, S.; Cosgrave, E.; Acuto, M.; Mcneill, D. Digital infrastructures and urban governance. *Urban Policy Res.* **2017**, *35*, 20–31. [CrossRef]
60. Ahmad, N.; Laplante, P.; DeFranco, J.; Kassab, M. A cybersecurity educated community. *IEEE Trans. Emerg. Top. Comput.* **2021**, *10*, 1456–1463. [CrossRef]

sustainability

Article

Synergy of Patent and Open-Source-Driven Sustainable Climate Governance under Green AI: A Case Study of TinyML

Tao Li [1,2,3], Jianqiang Luo [4], Kaitong Liang [1,2,*], Chaonan Yi [1,2] and Lei Ma [2,5,*]

1 School of Intellectual Property, Nanjing University of Science and Technology, No. 200 Xiaolingwei Street, Xuanwu District, Nanjing 210094, China; taolee@njust.edu.cn (T.L.); njlgdx10095@njust.edu.cn (C.Y.)
2 Centre for Innovation and Development, Nanjing University of Science and Technology, Nanjing 210094, China
3 School of Business, Xianda College of Economics & Humanities Shanghai International Studies University, No. 390 Dong Tiyuhui Rd., Hongkou District, Shanghai 200083, China
4 School of Management, Jiangsu University, No. 301 Xue Fu Rd., Zhenjiang 212013, China; ljq2809@163.com
5 School of Public Affairs, Nanjing University of Science and Technology, Nanjing 210094, China
* Correspondence: liangktong@njust.edu.cn (K.L.); maryma2009@njust.edu.cn (L.M.)

Abstract: Green AI (Artificial Intelligence) and digitalization facilitate the "Dual-Carbon" goal of low-carbon, high-quality economic development. Green AI is moving from "cloud" to "edge" devices like TinyML, which supports devices from cameras to wearables, offering low-power IoT computing. This study attempts to provide a conceptual update of climate and environmental policy in open synergy with proprietary and open-source TinyML technology, and to provide an industry collaborative and policy perspective on the issue, through using differential game models. The results show that patent and open source, as two types of TinyML innovation, can benefit a wide range of low-carbon industries and climate policy coordination. From the case of TinyML, we find that collaboration and sharing can lead to the implementation of green AI, reducing energy consumption and carbon emissions, and helping to fight climate change and protect the environment.

Keywords: climate governance; environmental sustainability; green AI; TinyML; patent; open source

Citation: Li, T.; Luo, J.; Liang, K.; Yi, C.; Ma, L. Synergy of Patent and Open-Source-Driven Sustainable Climate Governance under Green AI: A Case Study of TinyML. *Sustainability* **2023**, *15*, 13779. https://doi.org/10.3390/su151813779

Academic Editors: Tan Yigitcanlar, Bo Xia, Jamile Sabatini Marques and Tatiana Tucunduva Philippi Cortese

Received: 4 August 2023
Revised: 10 September 2023
Accepted: 13 September 2023
Published: 15 September 2023

1. Introduction

Major economies are regaining growth while facing challenges of carbon reduction and environmental protection [1]. Digital wave offers opportunities to achieve the goal of "Dual Carbon" [2], and green AI (Artificial Intelligence) is a tool designed to achieve quality development in a low-carbon economy [3]. As basic research on AI continues to progress [4–7], AI is accelerating from the "cloud" to the "edge" in the fields of smart manufacturing and smart cities [8], into smaller IoT devices for low-carbon energy-efficient computing and public information monitoring [9].

Negative impacts of AI on climate policy include increased electricity consumption and carbon emissions [10], i.e., threats such as cryptocurrencies. A shift to sustainable AI is imperative [3,11,12]. Inclusive, credible, explainable, ethical, and responsible technological approaches are required to drive smart city transformation [13] to mitigate planetary issues in a sustainable manner [13,14].

TinyML (Tiny Machine Learning) supports smart cameras, remote monitoring devices, wearable devices, audio capture hardware, and various sensors [15]. The power consumption and carbon footprint of TinyML devices are much lower than those of cloud computing and ordinary mobile devices, i.e., TinyML devices operate at a MHz level and consume power at a mW level, which is 1000 orders of magnitude lower than cloud computing and mobile devices; CO_2 emission levels are at a kg level, which is one order of magnitude lower.

Undoubtedly, an ecosystem of at least tens of billions of IoT devices will gain machine learning capabilities [16]. Low-carbon and green TinyML can create a healthier and more sustainable environment [17].

We utilize WIPO's PATENTSCOPE database to search for TinyML keywords in searchable fields to obtain relevant patent document information, and we then use crawler software to capture data. We manually referenced IPC code to group, classify, and count these TinyML patents. Second, we retrieve TinyML-related open-source projects from GitHub open-source code hosting platform and organize and classify these open-source projects' programming languages and count them manually. The data in Tables 1 and 2, both sorted in descending order, are obtained through the above acquisition and cleaning process. As the tables show, there are differences in the classification of patent and open source because patent classification focuses more on categorizing inventions as specific technological domains to demonstrate independence and systematic nature [18]. Open-source projects are usually artifacts that depend on each other and other components to build a fully functional system [19], whereas classification of open-source projects repositories is more inclined to functional areas so that developers target and contribute to specific problems or functions.

Table 1. TinyML Patents' IPC Code and Usage Classification Statistics based on WIPO's PATENTSCOPE.

IPC Code	IPC Usage Classification	Count
G06N	computing model	46
G06F	data processing	30
H04L	data transmission	24
G06Q	e-commerce and e-government	19
G06K	data visualization	18
G05B	control system	11
A61B	analytical biology	9
G16Y	IoT communications	8
B25J	robotic arm	6
G06V	computer vision	5
Grand Total		176

Data retrieved as of 1st May 2023.

Table 2. TinyML Open-Source Project Repositories and Artifact Classification Statistics in GitHub.

Programing Language	Artifact Classification	Count
Other (i.e., Assembly, Java, Arduino)	firmware	499
C++	hardware	201
C	hardware	131
Python	software	74
F#	software	14
JavaScript	GUI	10
Go	software	9
HTML	GUI	6
TeX	typesetting	5
Jupyter Notebook	data science	4
CSS	GUI	4
Grand Total		957

Data retrieved as of 1st May 2023.

WIPO patent data (Table 1) indicate that TinyML green AI technology has expected economic benefits in algorithm applications, data processing, control systems, IoT devices, etc. (Figure 1). Open-source data (Table 2) show that repositories related to TinyML in GitHub are mainly distributed in firmware, hardware programming, and algorithm implementation and application at the software level (Figure 2).

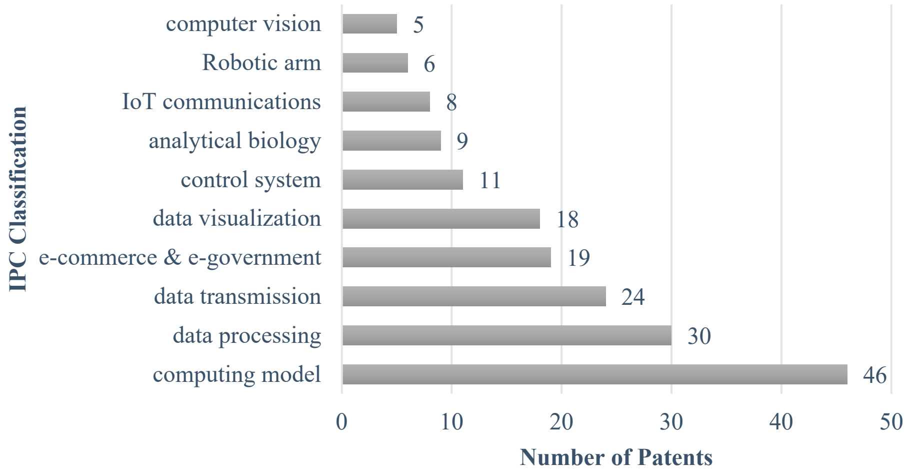


Figure 1. TinyML Green Technology in Patent Innovation Distribution.

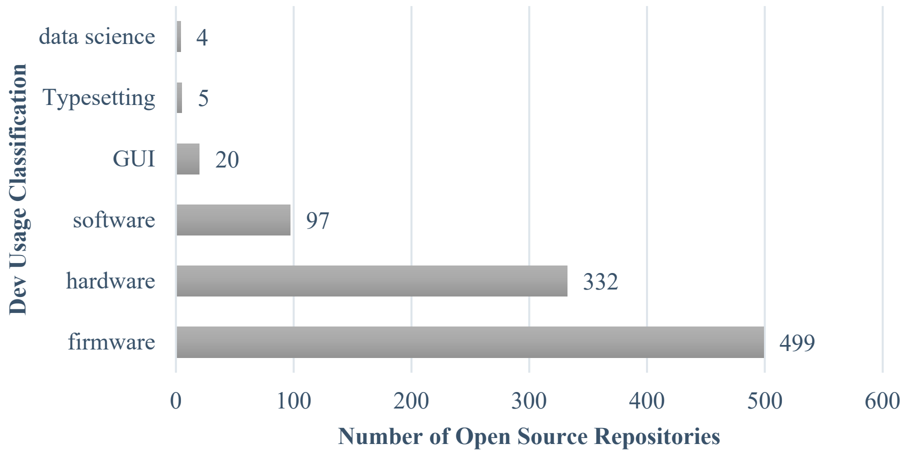


Figure 2. TinyML Green Technology in Open-Source Innovation Distribution.

TinyML's patented and open-source technologies play an essential role in climate policy [20]. Optimizing irrigation for precision agriculture is used to improve crop yield and quality [21,22]. Smart sensing systems are used in the field of early warning and monitoring systems [23,24] to address the challenges posed by natural disasters and climate change [25]. Intelligent energy-management systems reduce energy waste and carbon emissions [26]. In smart cities, optimizing energy use of appliances [27] improves energy efficiency [28] and reduces energy costs [29]. Optimizing traffic flow in intelligent transportation systems reduces traffic congestion and carbon emissions [30]. In conclusion, patented and open-source implementations of TinyML technology can play a positive role in the development of aspects of climate policy.

Patent and open source are two distinct forms of innovation. Knowledge spillovers bring these approaches to innovation closer together [31–33]. Organizations are blending these two approaches to gain advantages in competition [34,35]. Through open-source proactive spillover of technical knowledge, people draw inspiration from existing source code examples or leverage feedback from the community to innovate or improve products [36]. Meanwhile, patent-built barriers are used to protect inventions [37]. In order to form a TinyML green technology cluster, organizations must have both proprietary and open-source technologies. Proprietary technologies create a competitive advantage and bring economic benefits. And open-source technologies facilitate collaboration and expand value networks [38]. Through the synergy of proprietary and open-source technologies, a more efficient and productive cluster can be constructed [39,40].

Research on synergy of patent and open source in TinyML-based green AI is yet to be studied. We try to dissect this issue and also provide a broader vision of industry synergy and policy for the increasingly serious climate and energy challenges.

The significance of this study lies in the fact that it shows the impact of TinyML-based green AI with patent and open-source synergies on climate and environmental governance, and demonstrates the potential to improve efficiency of climate and environmental governance by providing technology-based conceptual updates for low-carbon, green, and sustainable climate policies.

The first section explains the research background, raises research questions, introduces methods, and describes the contributions. The second section discusses related works in recent years that support our study. In the third section, we introduce differential games, showing the results in three situations of noncooperative game, Stackelberg leader–follower game, and cooperative game, and conduct discussions. In the fourth section, we summarize the results of innovative synergy in the areas of proprietary and open source for TinyML low-carbon green technology, and extrapolate to a wider range of low-carbon industries and climate policy collaboration.

Realistic strategic decision-making is a game model that considers time factor and dynamic changes, and differential games provide an analytical approach [41]. Nash's noncooperative game [42] lays the theoretical foundation for differential games, which has become a critical branch of modern game theory and plays crucial role in analyzing competitive strategy, market behavior, and strategic decision-making in economics, management, and engineering [43]. In environmental protection and green AI, patents focus on securing independence of innovative achievements by obtaining economic returns through the granting of patent usage rights. Open source, on the other hand, emphasizes knowledge sharing and collaboration, and promotes the popularization and progress of technological innovation by disclosing the design and source code of innovative achievements. Through differential game analysis of competition conditions, behavioral patterns of participants, and strategic choices and dynamics of competition between patent and open-source approaches are revealed.

The scope of this study includes a conceptual update of climate and environmental policy to incorporate TinyML technology. The analysis covers the potential of TinyML to address low-carbon and climate policy issues through the examination of patent data and open-source data. By using differential game theory to model competition between patent and open-source approaches to environmental protection and low-carbon development, the scope extends to providing an industry collaborative and policy perspective on the issues, with a focus on collaboration, sharing, energy consumption reduction, and carbon emissions reduction for sustainable development in environmental protection.

First, this study makes a marginal contribution to interdisciplinary theories of environmental sustainability and digital innovation: data mining of patent and open source allows for conceptual updates of climate policy and reveals the potential of TinyML-based green AI in balancing environment and efficiency. Secondly, comparing previous works that focus solely on noncooperative [44] or Stackelberg [45] approaches, or on a closed-loop supply chain [46], we examine the application of differential game theory to evaluate complex competing synergistic mechanisms by introducing several key variables and parameters, i.e., rate of open technology value, willingness to open, impact coefficients, decay rates, long-term profits and benefits of combining noncooperative, Stackelberg, and cooperative games, and expanding to industrial competitive synergies.

2. Literature Review

Due to the rapid development of the digital era, smart technologies are seen as an effective tool for solving challenging issues facing the world today and mitigating environmental, social, and economic crises on a global scale [47]. The EU's "Green New Deal" sets out the strategic goal of decoupling economic growth from resource use [48,49]. Industry 4.0 has given rise to the new concept of Cities 4.0 [50], which aims to improve quality of life, productivity, and sustainability of cities with AI [51]. Sustainability profoundly influences the direction of energy, transportation, housing, and agriculture [52]. Green AI advocates a "circular economy" that aims to reduce, reuse, and recycle across sectors and geographies [53]. Thus, the research value of green AI is highlighted. Human capital, financing power, technological innovation, and government policies play critical roles in the green transformation of AI [54]. Patent and open source provide the technical knowledge required to integrate intelligence and greenness [55].

The essence of AI innovation lies in fulfilling efficient and accurate intelligent algorithms [56] to promote humanistic and responsible technological development [57] and social progress [58,59]. Through continuous exploration and improvement of green AI technologies to reduce dependence on natural resources, and multi-disciplinary cooperation and application, people are forging a sustainable [14] intelligent path for the future of human beings and the planet [60]. TinyML [61] implements efficient machine learning models on edge devices with low power consumption and low resource consumption [62], reduces energy dependence, lowers the burden on the environment, which drives green AI technologies, and redefines smart cities [13].

The application and promotion of TinyML is increasing [63]. In recent years, TinyML as a typical technology for green AI is continuously progressing in climate governance, environmental protection, precision agriculture, and smart cities, as demonstrated below.

2.1. In Environmental and Climate Governance

Reducing carbon footprint is crucial. TinyML as green AI is an important tool for realizing climate and environmental policies [29], e.g., overcoming limitations of traditional sensors and monitoring systems [20], superior efficiency and energy saving advantages over traditional machine learning algorithms when running on small devices [64], low power consumption, high efficiency and energy saving capabilities, and low storage costs in environmental radiation-monitoring systems [65,66]. It also offers adaptive unsupervised anomaly detection for extreme environments [67], thus facilitating provision of accurate data for weather forecasting and disaster warnings [68]. The combination of TinyML and CloudML (cloud machine learning) even enhances environmental monitoring and climate prediction [69].

2.2. In Precision Agriculture

TinyML-based green AI addresses key challenges in precision agriculture by providing better tuning of environmental parameters [70], reducing resource consumption [71], improving crop yield and quality [72], and promoting sustainable development. As an example, the TinyML intelligent control system outperforms traditional models in maintaining temperature and humidity balance [73], reducing system response time and resource consumption [74], achieving smarter and more efficient food production [75], reducing energy waste and environmental pollution [76], and thus protecting the environment and mitigating climate change [17].

2.3. In Smart Cities

Industry 4.0 has spawned the new concept of Cities 4.0 [50]. TinyML better realizes the collection and management of urban data. The Intelligent Transportation System (ITS), based on TinyML IoT, can reduce traffic congestion and pollution [71], and promote the green and low-carbon development of smart cities [47,51,77].

2.4. Technological Innovation

TinyML promotes technological innovation through algorithmic optimization. The TinyML algorithm improves energy output efficiency by implementing a Square Cross-section Two-phase Closed heat pump (SCTC) in a photovoltaic (PV) system [78]. TinyReptile, a decentralized edge machine learning model, combines TinyML and coalitional meta-learning to improve computational efficiency and performance [79]. Generalized TinyML benchmarking framework based on different operating system platforms lays the foundation for evaluation [80]. The TinyML compression algorithm reduces memory usage and computational complexity [81], enabling energy-efficient reasoning on Unmanned Aerial Vehicles (UAVs) [28]. The efficiency of the C4.5 decision tree algorithm is improved by determining economic granularity interval in TinyML algorithm optimization [82].

2.5. Competition and Synergy between Proprietary and Open Source

Differences between open source and proprietary have different impacts on the software industry [83] and can be strategically complementary [35] and balanced [84]. Open source and patents have strong synergies [85]. Patents often have a technological lead over open source, but open source can also compete effectively [86]. Patents use a lock-in strategy, while open source offers greater flexibility and freedom [87,88]. Open source is more reliable and secure due to open management and auditing of source code [89,90]. RIVICE (Open Source River Ice Model) demonstrates the benefits of open source in environmental research collaboration and problem solving [91]. Migration timing framework from patents to open source provides strategic guidance [53], and the governance model and platform ecosystem will change [92].

The literature referred to above shows that TinyML's patent and open-source technologies provide additional opportunities for climate governance and environmental protection in multiple areas. Synergy needs to be further investigated.

3. Methods

3.1. Question

In TinyML-based green AI, patent and open source are located at the two ends of the innovation spectrum, with patents bringing economic benefits by establishing technological barriers and open source expanding value networks through proactive knowledge spillovers [31–33], affecting policy on climate and carbon reduction [20]. Patent innovation and open-source innovation form an innovation loop, as shown in Figure 3. The following subsections investigate the open synergy problem of TinyML patent and open source in the context of climate change and carbon reduction through three scenarios: noncooperative game, Stackelberg leader–follower game, cooperative game.

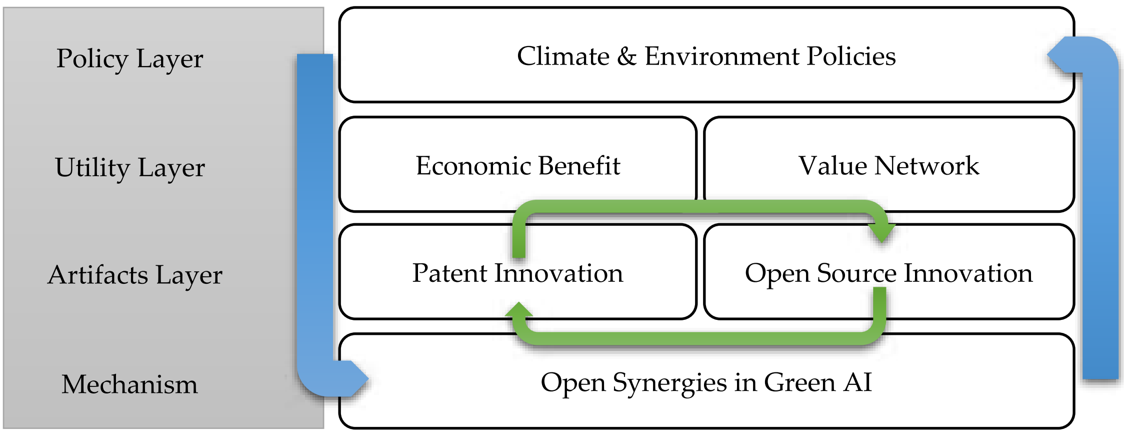


Figure 3. Green Technology Open Synergies for Climate and Environmental Policy Enhancement.

3.2. Premises

Underlying assumption: Proprietary and open-source platforms are participants in this study, both have technological achievements, conform to limited rationality, and make decisions with the goal of maximizing benefits [93]. The cost of technology openness is a convex function of willingness to open up technology [94].

PAPI components: To present differential games, we investigate the Players + Actions + Payoffs + Information (PAPI) framework [95], which reflects dynamic interactions between players' decisions based on the information they observe and the behaviors of others. Competitive synergy is a long-term process with participants' decisions adjusted over time, and they influence each other.

P(Players): Patent platform and open-source platform are participants in this study.

A(Actions): Both players have strategies set {Open, Close} in a discrete view. However, as differential games consider a multidimensional continuum of strategic changes and confrontations [96], players' strategies vary depending on value rate of opening technologies (denoted μ), willingness to open up technology (E), input costs to open up technologies (C), impact coefficient on open synergy effect (λ), decay rate of synergies (δ), marginal return (π), impact factor of technology opening on total return (θ), revenue sharing rate (α), degree of technology openness from patent to open source (σ), discount factor for both platforms (ρ). The relevant symbols are defined in Table 3.

Table 3. Symbols and Description.

Symbol	Description
X	Patent platform
Y	Open-source platform
$\mu_X \in [0,1]$	Value rate of opening up technologies on patent platform
$\mu_Y \in [0,1]$	Value rate of opening up technologies on open-source platform
$E_X(t) \in \mathbb{R}$	Willingness to open up technology on patent platform
$E_Y(t) \in \mathbb{R}$	Willingness to open up technology on open-source platform
$C_X(t) \in \mathbb{R}$	Input costs for patent platform to open up technologies
$C_Y(t) \in \mathbb{R}$	Input costs for open-source platform to open up technologies
$\lambda_X \in [0,1]$	Impact coefficient of patent platform on open synergy effect in TinyML technology
$\lambda_Y \in [0,1]$	Impact coefficient of open-source platform on open synergy effect in TinyML technology
$\delta \in [0,1]$	Decay rate of synergies due to technology opening
$K(t) \in \mathbb{R}$	Synergies from technology opening up at moment t
$\pi_X \in [0,1]$	Marginal return of patent platform
$\pi_Y \in [0,1]$	Marginal return of open-source platform
$\Pi(t) \in \mathbb{R}$	Total return from technology opening
$\theta \in [0,1]$	Impact factor of technology opening up on total return
$\alpha \in [0,1]$	Revenue sharing rate for patent platform
$1-\alpha \in [0,1]$	Revenue sharing rate for open source platform
$\sigma \in [0,1]$	Degree of technology openness from patent platform to open-source platform
$\rho \in [0,1]$	Discount factor for both platforms
$J_X \in \mathbb{R}$	Long-term profits of patent platform
$J_Y \in \mathbb{R}$	Long-term profits of open-source platform
$V_X(K) \in \mathbb{R}$	Benefits from technology opening up on patent platform
$V_Y(K) \in \mathbb{R}$	Benefits from technology opening up on open source platform

P(Payoffs): Payoffs of both players in differential games are measured in terms of their total return from respective strategies. For simplicity, we model interactions by providing a payoffs matrix (Table 4) under three differential game models: noncooperative game, Stackelberg leader–follower game, and cooperative game. The process of equilibrating the games based on HJB equations is given sequentially in Section 3.3.

Table 4. Payoffs under Differential Games.

	Payoffs for Patent Platform	Payoffs for Open-Source Platform
Noncooperative game	V_X^* (10)	V_Y^* (11)
Stackelberg leader–follower game	V_X^{**} (17)	V_Y^{**} (18)
Cooperative game	V_X^{***} (23)	V_Y^{***} (24)

Note: This is a matrix for preview model interactions by both players. Process of equilibrating the games based on HJB equations is given sequentially in Section 3.3.

I(Information):

The patent platform is denoted by X. The value rate of open green AI technology is μ_x. The willingness to open is a time-varying function $E_X(t)$. The input cost by patent platform is denoted as Equation (1).

$$C_X(t) = \frac{\mu_X}{2}[E_X(t)]^2 \tag{1}$$

The open-source platform is denoted by Y. The value rate of open green AI technology is μ_Y. The willingness to open is a time-varying function $E_Y(t)$. The input cost by open-source platform is denoted as Equation (2).

$$C_Y(t) = \frac{\mu_Y}{2}[E_Y(t)]^2 \tag{2}$$

Innovation and technology accumulation of green AI technologies in patent and open-source platforms have a positive incentive effect on technology opening, generating a synergistic effect of technology opening [97], enhancing innovation efficiency of smart green technology industry, and strengthening the policy effect of climate [34].

The synergy effect [98] is a time-varying function [99]. The rate of decay [100] describes decay of synergistic utility over time, given that other variables are equal. λ_X is the influence coefficient, which describes the open synergy effect of the patent platform on TinyML green technology. λ_Y is the influence coefficient, which describes the open synergy effect of the open-source platform on TinyML green technology. δ represents the decay rate of synergies from technology opening. The synergistic effect of technology opening up at time t is denoted as $K(t)$. TinyML's patent and open-source differential equations for generating synergies are as follows:

$$\frac{\partial}{\partial t}K(t) = \lambda_X E_X(t) + \lambda_Y E_Y(t) - \delta K(t) \tag{3}$$

The impact factor of technology openness on total return is denoted as θ. The marginal returns of patent and open-source platforms are π_X and π_Y, respectively. $\Pi(t)$ is the total return of technology openness in the following equation [101].

$$\Pi(t) = \pi_X E_X(t) + \pi_Y E_Y(t) + \theta K(t) \tag{4}$$

The total returns of technology opening are shared between platforms based on rules and mechanisms. The revenue share rate of the patent platform is denoted as $\alpha \in (0,1)$. The revenue share of the open-source platform is $1-\alpha$. Platforms incentivize each other to open up technology. The incentive level is denoted as $\sigma \in [0,1]$ for the patent platform and $1-\sigma$ for the open-source platform. The discount factor is assumed to be positive for both platforms and is denoted as ρ.

Long-term profit functions for two platforms are expressed as follows.

$$\begin{cases} J_X = \int_0^\infty e^{-\rho t}[\alpha\Pi(t) - C_X(t) - \sigma C_Y(t)]dt \\ J_Y = \int_0^\infty e^{-\rho t}[(1-\alpha)\Pi(t) - (1-\sigma)C_Y(t)]dt \end{cases} \tag{5}$$

Symbols used in this section and meanings are listed in Table 3.

The following subsections discuss the results of the two platforms in three game scenarios: noncooperative game, Stackelberg leader–follower game, and cooperative game.

3.3. Game Model

3.3.1. Noncooperative Game

The noncooperative game is the one in which each player has its own interests and goals, and where players do not necessarily need to cooperate with each other to achieve each optimal interest. The patent platform and open-source platform were first regarded as a noncooperative game. Micro and small participants in the patent platform own patented technologies and license or trade them to gain financial benefits. In the open-source platform, developers use and share source codes for free. The open environment creates higher visibility and a greater social effect. Everyone can contribute knowledge and skills, and communicate and collaborate across communities.

In the noncooperative game of the patent and open-source platforms, players make their own decisions and actions to achieve their own optimal interests. In the gaming process, the Nash equilibrium allows us to understand the relationship between cooperation and competition and to find equilibriums.

In the noncooperative game, platforms aim to maximize respective profits and benefits from each technology opening and are a function of synergistic effects, denoted as $V_X(K)$ and $V_Y(K)$, respectively, and satisfy the HJB equation.

$$\rho V_X(K) = \max_{E_X \geq 0}\left[\alpha\Pi(t) - C_X(t) + {V_X}'(k)\frac{\partial}{\partial t}K(t)\right] \tag{6}$$

$$\rho V_Y(K) = \max_{E_Y \geq 0}\left[(1-\alpha)\Pi(t) - C_Y(t) + {V_Y}'(k)\frac{\partial}{\partial t}K(t)\right] \tag{7}$$

By solving the HJB equation, we obtain optimal willingness of each platform to open up its technology and optimal total revenue of each.

$$E_X^* = \frac{\alpha(\pi_X(\rho+\delta) + \theta\lambda_X)}{\mu_X(\rho+\delta)} \tag{8}$$

$$E_Y^* = \frac{(1-\alpha)(\pi_Y(\rho+\delta) + \theta\lambda_Y)}{\mu_Y(\rho+\delta)} \tag{9}$$

$$V_X^*(K) = \frac{\alpha\theta}{\rho+\delta}K + \frac{\alpha^2\mu_X}{2\rho}(E_X^*)^2 + \frac{\alpha\mu_Y}{(1-\alpha)\rho}(E_Y^*)^2 \tag{10}$$

$$V_Y^*(K) = \frac{(1-\alpha)\theta}{\rho+\delta}K + \frac{\alpha\mu_X}{(1-\alpha)}(E_X^*)^2 + \frac{\mu_Y}{2\rho}(E_Y^*)^2 \tag{11}$$

The optimal total benefit (return) is the sum of $V_X^*(K)$ and $V_Y^*(K)$.

$$\begin{aligned} V^*(K) = V_X^*(K) \quad & +V_Y^*(K) \\ & = \tfrac{\theta}{\rho+\delta}K + \tfrac{\alpha(2-\alpha)(\pi_X(\rho+\delta)+\theta\lambda_X)^2}{2\rho\mu_X(\rho+\delta)^2} \\ & + \tfrac{(1-\alpha^2)(\pi_Y(\rho+\delta)+\theta\lambda_Y)^2}{2\rho\mu_Y(\rho+\delta)^2} \end{aligned} \tag{12}$$

3.3.2. Stackelberg Leader–Follower Game

The Stackelberg leader–follower game is noncooperative game model in which players are composed of a leader and a subordinate (or follower). The leader makes the first decision, and the follower makes decisions based on the leader's decision. The leader pre-observes the response of the follower and then makes an optimal decision, while the subordinate needs to follow the leader's decision to make a response. Thus, the leader has the advantage of formulating the best strategy, while the subordinate needs to react to the leader's strategy to achieve the best response.

There is a relationship between technology sharing and competitiveness, and the Stackelberg model helps scholars study the impact of technology sharing on strategies and final market competitiveness [102]. Moreover, part of the Stackelberg game involves technical open sharing between a patent and an open-source platform. Patent holders obtain informational and technological advantages by applying for patents and, therefore, utilize this advantage to formulate optimal strategies. The patent platform effectively controls market pricing and market entry barriers, thus gaining higher profits. The open-source platform, on the other hand, needs to respond to the strategies of the patent platform to maximize resources and technological advantages. The open-source platform utilizes free and open attributes to attract developers and users, providing better user experience and higher-quality products. However, open-source platforms need to comply with patent and intellectual property regulations, and sometimes attempt to circumvent patents [103]. In addition, open-source platforms need to continually innovate and expand the user base to maintain competitive advantages.

In practice, both open-source and patent platforms continually adjust their strategies based on changes in market demand and technological progress. In the Stackelberg leader–follower game, the patent platform acts as leader, while the open-source platform acts as subordinate. Both parties need to continually evaluate market demand and technological trends and formulate corresponding strategies to gain maximum benefits. Assume the degree of technology openness of the patent platform to the open-source platform is $\sigma \in [0,1]$ and the open-source platform follows the decision based on the level of technologies it already possesses.

σ represents the degree of technology openness of the patent platform to the open-source platform. σ takes a value of [0,1], which represents a continuum from no technology openness at all ($\sigma = 0$, no technology is shared) to complete openness ($\sigma = 1$, all technologies are openly shared).

When $\sigma = 0$, patent platform does not open any technology to the open-source platform at all, which means the patent platform completely retains its own technology secrets and does not share any technology with the open-source platform.

When $\sigma = 1$, the patent platform completely opens all technologies to the open-source platform, which means the patent platform is willing to share all technologies with the open-source platform and does not keep any secrets of its technologies.

Benefit from technology openness is a function of synergistic effect from technology openness, denoted as $V_X(K)$ and $V_Y(K)$, respectively, and satisfies the HJB equation. Optimal control of the open-source platform as a follower is described in Equation (13).

$$\rho V_Y(X) = \max_{E_Y \geq 0}\left[(1-\alpha)\Pi(t) - \frac{\mu_X}{2}(1-\sigma)E_Y^2 + V_Y{}'(K)\frac{\partial K(t)}{\partial t}\right] \tag{13}$$

Optimal technology openness intentions for the patent and open-source platforms are as follows.

$$E_X^{**} = \frac{\alpha[(\rho+\delta)\pi_X + \theta\lambda_X]}{(\rho+\delta)\mu_X} \tag{14}$$

$$E_Y^{**} = \frac{(1+\alpha)[(\rho+\delta)\pi_Y + \theta\lambda_Y]}{2\mu_Y(\rho+\delta)} \tag{15}$$

Optimal incentive levels for the patent and open-source platforms are as follows.

$$\sigma^{**} = \begin{cases} \frac{2\alpha-1}{1+\alpha}, \alpha \in (1/3, 1] \\ 0, \alpha \in (0, 1/3] \end{cases} \tag{16}$$

Optimal and total returns for the patent and open-source platforms are as follows.

$$V_X^{**}(K) = \frac{\alpha\theta}{\rho+\delta}K + \frac{\alpha^2[(\rho+\delta)\pi_X + \theta\lambda_X]^2}{2\rho\mu_X(\rho+\delta)^2} + \frac{(1+\alpha)^2[(\rho+\delta)\pi_Y + \theta\lambda_Y]^2}{8\rho\mu_Y(\rho+\delta)^2} \tag{17}$$

$$V_Y^{**}(K) = \frac{(1-\alpha)\theta}{\rho+\delta}K + \frac{\alpha(1-\alpha)[(\rho+\delta)\pi_X + \theta\lambda_X]^2}{\rho\mu_X(\rho+\delta)^2} + \frac{\left(1-\alpha^2\right)[(\rho+\delta)\pi_Y + \theta\lambda_Y]^2}{4\rho\mu_Y(\rho+\delta)^2} \tag{18}$$

$$V^{**}(K) = V_X^{**}(K) + V_Y^{**}(K) = \frac{\theta}{\rho+\delta}K + \frac{\alpha(1-\alpha/2)[(\rho+\delta)\pi_X + \theta\lambda_X]^2}{\rho\mu_X(\rho+\delta)^2} + \frac{\frac{1}{2}\left(3-2\alpha+3\alpha^2\right)[(\rho+\delta)\pi_Y + \theta\lambda_Y]^2}{\rho\mu_Y(\rho+\delta)^2} \tag{19}$$

Using the patent platform as leader, as described above, can be extrapolated to, e.g., a small patent firm (SPF) or a small open-source firm (SOSF). (1) If the SPF acts as leader, it seeks patent protection more aggressively to ensure exclusivity and market dominance of its innovations in order to maintain inflow of economic benefits and operational sustainability. The SPF adopts more conservative strategies to control the release and sharing of intellectual property to protect its interests. (2) If the SOSF acts as leader, it should emphasize open-source innovation and collaboration, and share knowledge and source code with other developers and researchers. This promotes technological progress and market development while increasing influence and competitiveness. Open-source developing lends itself to large-scale collaboration and sharing, thus posing a significant challenge to the SOSF with limited resources. In brief, small firms tend to face constraints in terms of resources and size in highly competitive markets. Industrial centrality of innovations, flexibility of innovation, and diversity of innovation cooperation networks are crucial for becoming a leader. Both open-source and patent small firms need appropriate strategies based on resources and goals, and adapt to a competitive environment and market changes.

3.3.3. Cooperative Game

The cooperative game is one in which players make decisions with the goal of maximizing common interests. Sometimes, all players maximize their gains, while at other times, one player needs to lose its benefits to maximize collective benefits, thus maximizing sustainability of cooperation. A cooperative game of technology openness is a way to achieve a mutually beneficial situation through open technology. In the technology openness scenario, various players cooperate through technology sharing and knowledge transfer [104], i.e., patent platform facilitates exchange data and innovation through open data sharing. Data providers and users collaborate to achieve appropriate data usage. Patent owners use the open-source platform to allow more users to use and submit contributions, increasing the stability and functionality of the technology, as well.

The patent platform and open-source platform allocate total benefits obtained from the cooperative game of technology opening at a rate of $\alpha \in [0, 1]$ for the patent platform and $1 - \alpha$ for the open-source platform. Revenues as total benefits received by the patent platform and open-source platform are based on technology openness allocation. When $\alpha = 0$, the patent platform receives zero benefits and all benefits are retained in the open-source platform, which means that the patent platform fully opens up its technology and is willing to give up its own financial interests for the development of an open-source community. When $\alpha = 1$, the open-source platform gains zero benefits and all benefits go to the patent platform, which means that the patent platform fully retains revenues and shares no technology with the open-source platform. The value of α can be determined according to the interests and goals of both parties to the cooperation, and it is an adjustable parameter that can regulate the distribution of benefits in the cooperation agreement.

Benefits of technology opening for both players are denoted as $V_X(K)$ and $V_Y(K)$, and total benefits are denoted as $V(K)$, and both satisfy the HJB equation.

$$\rho V(K) = \max_{E_X \geq 0, E_Y \geq 0} \left[\Pi(t) - \frac{\mu_X}{2} E_X^2 - \frac{\mu_Y}{2} E_Y^2 + V'(K) \frac{\partial}{\partial t} K(t) \right] \tag{20}$$

Open intentions for the patent and open-source platforms are as follows.

$$E_X^{***} = \frac{(\rho + \delta)\pi_X + \theta\lambda_X}{(\rho + \delta)\mu_X} \tag{21}$$

$$E_Y^{***} = \frac{(\rho + \delta)\pi_Y + \theta\lambda_Y}{(\rho + \delta)\mu_Y} \tag{22}$$

Optimal returns and optimal total return from open technology for both players are as follows.

$$V_X^{***} = \frac{\alpha\theta}{\rho+\delta}K + \frac{\alpha[(\rho+\delta)\pi_X + \theta\lambda_X]^2}{2\rho\mu_X(\rho+\delta)^2} + \frac{\alpha[(\rho+\delta)\pi_X + \theta\lambda_Y]^2}{2\rho\mu_Y(\rho+\delta)^2} \tag{23}$$

$$V_Y^{***} = \frac{(1-\alpha)\theta}{\rho+\delta}K + \frac{(1-\alpha)[(\rho+\delta)\pi_X + \theta\lambda_X]^2}{2\rho\mu_X(\rho+\delta)^2} + \frac{(1-\alpha)[(\rho+\delta)\pi_X + \theta\lambda_Y]^2}{2\rho\mu_Y(\rho+\delta)^2} \tag{24}$$

$$V^{***} = \frac{\theta}{\rho+\delta}K + \frac{[(\rho+\delta)\pi_Y]^2}{2\rho\mu_X(\rho+\delta)^2} + \frac{[(\rho+\delta)\pi_Y + \theta\lambda_Y]^2}{2\rho\mu_Y(\rho+\delta)^2} \tag{25}$$

The cooperative game on technology open scenarios is an important method of cooperation between enterprises, which promotes technological innovation and market competitiveness. Through the cooperative game of technology openness, enterprises mutually promote each other, develop together, and form a strong joint competitiveness, and also make positive contributions to the industry. The following section discusses the noncooperative game, Stackelberg leader–follower game, and cooperative game between patent and open source.

4. Discussion

4.1. Results Analysis

A numerical example is provided based on definitional and value domains in Table 1 in Section 3.2, by randomly selecting values for these variables and parameters:

$\alpha = 0.4, \theta = 0.3, \delta = 0.2, \rho = 0.2, \mu_X = 0.3, \mu_Y = 0.4, \lambda_X = 0.2, \lambda_Y = 0.3, \pi_X = 0.5$, $\pi_Y = 0.6$, and $K(t=0) = 2$, outcomes on willingness to open up technology (denoted E) for both players separately are obtained as $E_X^* \approx 0.8667 \leq E_X^{**} \approx 0.8667 < E_X^{***} \approx 2.1667$, and $E_Y^* \approx 1.2375 < E_Y^{**} \approx 1.4438 < E_Y^{***} \approx 2.0625$. It can be seen that the willingness to open up of both players increases progressively under the three types of games.

Further, benefits from technology opening (denoted V) for both players separately are obtained as $V_X^* \approx 3.20521 < V_X^{**} \approx 3.24775 < V_X^{***} \approx 3.7099$, and $V_Y^* \approx 4.12141 < V_Y^{**} \approx 4.37664 > V_Y^{***} \approx 4.29734$. It can be seen that the benefit of the patent platform is strictly partial order with Pareto improvement under three games. Benefits of the open-source platform are not strictly partial order, and the open-source platform has the highest benefit (denoted V_Y^{**}) as the follower role under the Stackelberg leader-follower game.

In addition, from patent and open-source platforms' synergistic view, the total benefit (return) denoted as V is obtained as $V^* \approx 7.32661 < V^{**} \approx 7.62439 < V^{***} \approx 8.00724$, which is strictly partial order with Pareto improvement.

The results and numerical data above show that the total benefits of both platforms are Pareto improvement under the three games and having Pareto optimality under synergy (Figure 4). The platform's intention to cooperate under the Stackelberg leader–follower game increases with the level of incentive for technology openness input, which is related to the design of the incentive mechanism. Both parties have a higher willingness to open up technology under the cooperative game than under the noncooperative game. Attitudes and behaviors towards technology openness are not stable but can be influenced by various factors. Intention to open up technology is higher under the cooperative game than the noncooperative game, which indicates that mutual collaboration and exchange tend to foster innovation, resulting in better technology and higher returns [105]. Therefore, when making technology opening decisions, companies should take these influencing factors into account, combine them with the actual situation, and make a more effective technology opening strategy to achieve better competitive advantage and business returns.

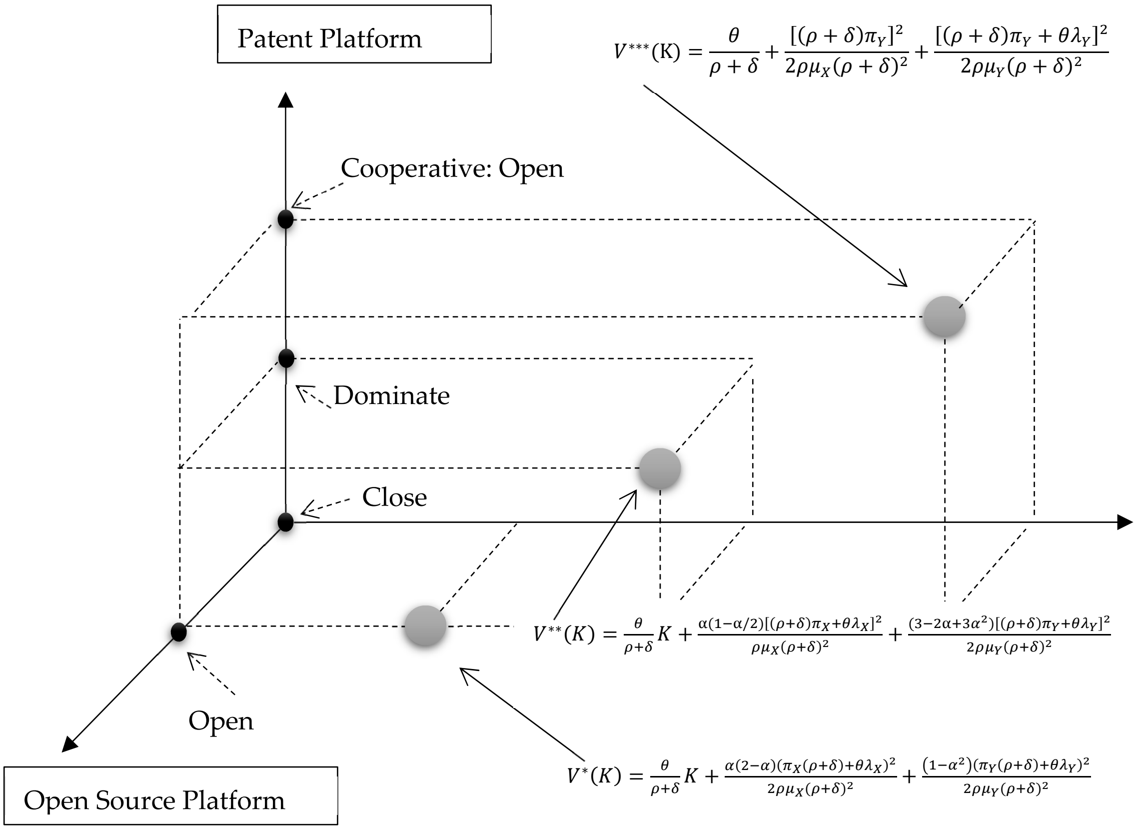


Figure 4. Level of total return for different strategy combinations.

(1) In the noncooperative game, in addition to considering changes in the willingness to open up technology, it is also necessary to consider the conflict of interest and incentives for noncooperation of the parties, including the effects of competition and uncertainty in the external market, resource allocation, and risk taking [106]. Strategies that can reduce incentives for noncooperation and conflicts of interest are formulated to achieve a better competitive advantage.

(2) In the Stackelberg leader–follower game, in addition to the level of incentives for technology openness, it is also necessary to analyze how much influence and control the leader has over the follower, as well as how much the follower trusts the leader and how receptive they are to technology openness decisions [107]. As mentioned in Section 3.3.2, role swapping between leader and follower is also an aspect worth investigating, i.e., when open-source platforms have sufficient core technology resources, the follower becomes the leader and has a dominant influence on the technology opening of the industry cluster [108]. Open-source platforms need to invest resources and incentives in order to encourage more developers and enterprises to join in technology sharing and innovation [109]. Meanwhile, the design of incentives for open-source platforms also needs to take into account the specificity and needs of the industry to establish a more flexible and effective mechanism for technological innovation [110]. Therefore, exploring the influence of power structure and role transformation on the willingness of technology openness in the Stackelberg leader–follower game is of great significance for enterprises when formulating technology openness strategies.

(3) In the cooperative game, strong willingness to open up technology and synergy between patent and open source has a boosting effect on innovation and returns. Sharing can bring mutual trust and cooperation, which leads to a greater market share and improved market competitiveness [111]. In the cooperative game, both players have higher trust and willingness to cooperate, which provides a more positive environment for technology opening. In addition, cooperation can also provide

more sharing of resources and expertise, which further promotes the development and innovation of technology [112]. Therefore, when enterprises make decisions on technology opening, they can consider establishing a stable cooperative game model through partnership, cooperative research and development, or technology sharing in order to achieve better results of technology opening [113].

The analysis above is insightful. Firms and policymakers should recognize the benefits of technology sharing and synergy in green AI industry. Firms should promote technology openness and incentivize collaboration while also protecting their own intellectual property rights. Policymakers should develop flexible laws and policies that encourage technology sharing and innovation while also promoting industrial upgrading and economic growth. Additionally, enterprises should strive for balance and caution in technology openness to protect their core technologies while engaging in healthy competition and collaboration with others. By adopting synergy, the TinyML-based green AI industry can achieve sustainable development and competitive advantages.

4.2. Synergistic Approach

Current climate and environmental governance policies should seek to actively engage with both patent and open-source areas of TinyML, promoting technology openness and innovative synergies between platforms. The following approaches are suggested.

(1) Promote the collaborative development of proprietary and open-source TinyML for climate and environmental governance goals. Use patent and open source together to develop TinyML components. This approach combines the advantages of patent and open source to ensure that TinyML development aligns with climate and environmental goals while remaining competitive and adaptable. Open patenting, a model that emphasizes knowledge sharing and collaboration, allows for technology sharing and licensing, fostering innovation and industrial progress [114]. Open patents can be a collective intelligence solution to drive technological innovation and adoption.

Under TinyML's climate and environmental governance objectives, open patents and open source can play a synergistic role in promoting the collaborative development of proprietary and open-source technologies [115].

In terms of open patent usage, proprietary technology companies can open up some of their patents to allow the open-source community to use them in their own projects, which can allow the open-source community to gain access to the application and technology experience of proprietary technology companies and build on it for better innovation.

In terms of building open-source frameworks, know-how companies can release their own products based on open-source frameworks, i.e., the open-source community, which can allow more developers to participate in the development of that product and ultimately create more solutions.

In terms of promoting co-development: proprietary technology companies and open-source communities can work together to develop new TinyML applications and share technology through open source, promoting continuous innovation in technology while learning from each other's experiences.

In terms of establishing a technology sharing platform, know-how companies can provide effective support for the open-source community in terms of know-how through a technology sharing platform with a focus on exchanging technology and reaching consensus on sharing technology experience, intellectual property, and patents, thereby creating more business opportunities.

(2) Foster co-promotion of proprietary and open-source TinyML under climate and environmental governance objectives. Proprietary and open-source platforms can work together to promote the use of TinyML technology. Patent and open-source platforms can collaborate to promote the use of TinyML technology. Open-source platforms can raise awareness and popularity by involving more developers in open-source hardware and software. Proprietary technologies can be applied to a wider range of fields through licensing agreements [116]. Collaboration among enterprises, organizations, and industry

groups can drive the development and application of TinyML technology in various sectors like smart agriculture and environmental protection. Strengthening the open-source community can enhance the dissemination and promotion of open-source technology. Active participation in policy and standard formulation can further enhance the role of TinyML technology in environmental governance and protection. Leveraging media, conferences, and forums can increase public awareness and understanding, promoting the development and application of TinyML technology in environmental governance and protection.

(3) Facilitate cross-fertilization of proprietary and open-source TinyML in the context of climate and environmental governance objectives. Patent and open-source technologies can collaborate to enhance the quality and efficiency of TinyML technology [117]. Open innovation can enable the joint development of new applications, allowing for knowledge exchange and improvement of technical expertise and effectiveness. Technical exchanges through conferences and forums can facilitate sharing of experiences and cases to advance TinyML. Establishing a talent training system can provide excellent professionals for both the open-source community and proprietary technology enterprises, promoting technology development and application. Collaboration in standardization efforts can better promote and implement TinyML standards. Creating a sharing platform for small, low-cost, and low-power TinyML devices can enable practitioners to share their experiences and learn from each other, enhancing the overall effectiveness of the technology.

In summary, promoting the mutual learning of proprietary and open-source TinyML in the context of climate and environmental governance objectives requires the open-source community and proprietary technology companies to work together to promote technology development and application, and to make a concerted effort in open innovation, technology exchange, talent training, standardization, and common sharing. In these ways, patented and open-source technologies can synergize and encourage each other to higher achievements and better results in the development of TinyML technology [88].

5. Conclusions

TinyML is machine learning based on small machine learning algorithms, low power consumption, low cost, and small data sets. Compared to conventional machine learning algorithms, TinyML's algorithms are designed to run on low-power devices such as IoT, portable devices, and drones. However, there may be competition between the two models of TinyML, proprietary and open source, which could be detrimental to the low-power IoT industry and climate environmental policy. In particular, in terms of patents, if technological innovation is blocked, then the industry will be stuck with a monopoly of knowledge and an inability to innovate. Especially in a highly specialized field such as TinyML, whether patents can be managed properly will directly affect industry costs, market competition, and the speed of industry innovation. The sustainability of the industry's development may also be affected if it relies only on open source.

As open source and patent technologies spread, this double-sided character intensifies. Which is more critical for green AI, open source or patented innovation roles? Open source freely discloses source code to the public to boost collaboration and joint improvement. In contrast, patent protects intellectual property and exclusive ownership. In the context of green AI, i.e., TinyML, where the focus should be on environmental friendliness, compared to high-energy consumption and pollution of traditional AI, which means both should contribute to the development and deployment of green AI solutions. Open-source innovation can foster collaboration and knowledge sharing among researchers and developers, leading to faster and wider development of AI technologies with environmentally friendly properties. Through open source, developers can improve on existing solutions, increasing efficiency and reducing environmental impact. On the other hand, patent innovation can encourage companies in green AI technological investment. Patents offer the right to protect inventions and realize innovations, which can drive commercialization and scaling of green AI solutions. Ultimately, open source and patent strike a balance. Open source can drive initial research and development of green AI, while patents can provide companies with the financial incentives necessary to bring technologies to market at scale. In brief, while open source and patent are critical in any technological advancement, they should be balanced with a focus on environmentally friendly use in green AI. A combination of open-source collaboration and patent protection can drive development and deployment of sustainable AI solutions that mitigate the effects of climate change.

First, this study makes a marginal contribution to interdisciplinary theories of environmental sustainability and digital innovation which provides a conceptual update of climate policy with the potential for TinyML-based green AI to balance environment and efficiency; second, this study contributes by applying differential game theory to evaluate complex competing synergistic mechanisms by introducing several key variables and parameters with combining noncooperative, Stackelberg, and cooperative games and expanding to industrial competitive synergies.

This study's limitations lie in a relatively novel direction with lag in disclosure of practice and research. Furthermore, as climate and environmental research is often related to ethical and sustainability issues, limitations such as data privacy, energy consumption, and environmental protection need to be carefully considered when applying TinyML technology.

Author Contributions: Conceptualization, T.L.; Data curation, T.L.; Formal analysis, T.L.; Funding acquisition, L.M.; Investigation, T.L. and L.M.; Methodology, T.L.; Project administration, L.M.; Resources, T.L.; Software, T.L.; Supervision, J.L. and L.M.; Validation, T.L., J.L., K.L. and C.Y.; Visualization, T.L.; Writing—original draft, T.L.; Writing—review and editing, T.L., J.L., K.L., C.Y. and L.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: All data and materials pertaining to this work have already been included within this manuscript in the form of either figures or tables.

Acknowledgments: Special thanks to Yun Jinhyo Joseph from JOI for invaluable advice provided during the SOItmc 2023 Naples Conference for improvement of our paper. This research has been supported by the following: (1) Major Project of National Social Science Foundation of China (NSSFC), "Research on the Mechanism and Path about Technological Standard and Intellectual Property Synergistically Promoting Digital Industry Innovation" (19ZDA078). (2) Jiangsu Decision-Making and Consultation Institution of Service-Oriented Government Development Fund from Nanjing University of Science and Technology. (3) Research Project of National Academy of Innovation Strategy (NAIS) "Research on the application of key technologies in digital economy at home and abroad" (2022-sjzx-02). (4) Research Project of 2022 Guangxi Philosophy and Social Sciences Planning Study "Mechanisms and Countermeasures for Deep Integration of Key Industry Chains and Innovation Chains in Guangxi towards ASEAN" (22BJY005). (5) ChenGuang Plan (Q99010.19.008), Xianda College Teaching and Research Division, Xianda College of Economics & Humanities Shanghai International Studies University. (6) 2022 Shanghai Municipality Special Funds for Promotion on Privately-run Education Development (Z30001.22.809), Xianda College of Economics & Humanities Shanghai International Studies University.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chong, C.T.; Fan, Y.V.; Lee, C.T.; Klemeš, J.J. Post COVID-19 ENERGY sustainability and carbon emissions neutrality. *Energy* **2022**, *241*, 122801. [CrossRef] [PubMed]
2. Ren, S.; Hao, Y.; Wu, H. Digitalization and environment governance: Does internet development reduce environmental pollution? *J. Environ. Plan. Man.* **2023**, *66*, 1533–1562. [CrossRef]
3. Schwartz, R.; Dodge, J.; Smith, N.A.; Etzioni, O. Green AI. *Commun. ACM* **2020**, *63*, 54–63. [CrossRef]
4. Li, P.; Peng, X.; Xu, C.; Han, L.; Shi, S. Novel extended mixed controller design for bifurcation control of fractional-order Myc/E2F/miR-17-92 network model concerning delay. *Math. Methods Appl. Sci.* **2023**, 1–21. [CrossRef]
5. Huang, C.; Wang, J.; Chen, X.; Cao, J. Bifurcations in a fractional-order BAM neural network with four different delays. *Neural Netw.* **2021**, *141*, 344–354. [CrossRef] [PubMed]

6. Xu, C.; Cui, Q.; Liu, Z.; Pan, Y.; Cui, X.; Ou, W.; Rahman, M.; Farman, M.; Ahmad, S.; Zeb, A. Extended Hybrid Controller Design of Bifurcation in a Delayed Chemostat Model. *Match Commun. Math. Comput. Chem.* **2023**, *90*, 609–648. [CrossRef]
7. Ali, Z.; Rabiei, F.; Hosseini, K. A fractal–fractional-order modified Predator–Prey mathematical model with immigrations. *Math. Comput. Simul.* **2023**, *207*, 466–481. [CrossRef]
8. Satyanarayanan, M. How we created edge computing. *Nat. Electron.* **2019**, *2*, 42. [CrossRef]
9. Arowolo, M.O.; Ogundokun, R.O.; Misra, S.; Agboola, B.D.; Gupta, B. Machine learning-based IoT system for COVID-19 epidemics. *Computing* **2023**, *105*, 831–847. [CrossRef]
10. Jin, L.; Duan, K.; Tang, X. What Is The Relationship between Technological Innovation and Energy Consumption? Empirical Analysis Based on Provincial Panel Data From China. *Sustainability* **2018**, *10*, 145. [CrossRef]
11. Dhar, P. The carbon impact of artificial intelligence. *Nat. Mach. Intell.* **2020**, *2*, 423–425. [CrossRef]
12. Song, M.; Cao, S.; Wang, S. The impact of knowledge trade on sustainable development and environment-biased technical progress. *Technol. Forecast. Soc.* **2019**, *144*, 512–523. [CrossRef]
13. Yigitcanlar, T.; Foth, M.; Kamruzzaman, M. Towards Post-Anthropocentric Cities: Reconceptualizing Smart Cities to Evade Urban Ecocide. *J. Urban. Technol.* **2019**, *26*, 147–152. [CrossRef]
14. Yigitcanlar, T.; Cugurullo, F. The Sustainability of Artificial Intelligence: An Urbanistic Viewpoint from the Lens of Smart and Sustainable Cities. *Sustainability* **2020**, *12*, 8548. [CrossRef]
15. Rajapakse, V.; Karunanayake, I.; Ahmed, N. Intelligence at the Extreme Edge: A Survey on Reformable TinyML. *ACM Comput. Surv.* **2023**, *55*, 1–30. [CrossRef]
16. Manavalan, M. Intersection of Artificial Intelligence, Machine Learning, and Internet of Things—An Economic Overview. *Glob. Discl. Econ. Bus.* **2020**, *9*, 119–128. [CrossRef]
17. Vuppalapati, C.; Ilapakurti, A.; Kedari, S.; Vuppalapati, J.; Kedari, S.; Vuppalapati, R. Democratization of AI, Albeit Constrained IoT Devices Tiny ML, for Creating a Sustainable Food Future. In Proceedings of the 2020 3rd International Conference on Information and Computer Technologies (ICICT), San Jose, CA, USA, 9–12 March 2020; pp. 525–530.
18. Benson, C.L.; Magee, C.L. Technology structural implications from the extension of a patent search method. *Scientometrics* **2015**, *102*, 1965–1985. [CrossRef]
19. Ma, Y.; Fakhoury, S.; Christensen, M.; Arnaoudova, V.; Zogaan, W.; Mirakhorli, M. Automatic Classification of Software Artifacts in Open-Source Applications. In Proceedings of the 15th International Conference on Mining Software Repositories, Gothenburg, Sweden, 28–29 May 2018; pp. 414–425.
20. Bamoumen, H.; Temouden, A.; Benamar, N.; Chtouki, Y. How TinyML Can be Leveraged to Solve Environmental Problems: A Survey. In Proceedings of the 2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Virtual, 20–21 November 2022; pp. 338–343.
21. Nicolas, C.; Naila, B.; Amar, R. TinyML Smart Sensor for Energy Saving in Internet of Things Precision Agriculture platform. In Proceedings of the 2022 Thirteenth International Conference on Ubiquitous and Future Networks (ICUFN), Barcelona, Spain, 5–8 July 2022; pp. 256–259.
22. Vuppalapati, C.; Ilapakurti, A.; Kedari, S.; Vuppalapati, R.; Vuppalapati, J.; Kedari, S. Crossing the Artificial Intelligence (AI) Chasm, Albeit Using Constrained IoT Edges and Tiny ML, for Creating a Sustainable Food Future. In Proceedings of the Fifth International Congress on Information and Communication Technology, London, UK, 20–21 February 2020; pp. 540–553.
23. Ronoh, E.K.; Mirau, S.; Dida, M.A. Human-Wildlife Conflict Early Warning System Using the Internet of Things and Short Message Service. *Eng. Technol. Appl. Sci. Res.* **2022**, *12*, 8273–8277. [CrossRef]
24. Tsoukas, V.; Gkogkidis, A.; Boumpa, E.; Papafotikas, S.; Kakarountas, A. A Gas Leakage Detection Device Based on the Technology of TinyML. *Technologies* **2023**, *11*, 45. [CrossRef]
25. Abdalzaher, M.S.; Elsayed, H.A.; Fouda, M.M.; Salim, M.M. Employing Machine Learning and IoT for Earthquake Early Warning System in Smart Cities. *Energies* **2023**, *16*, 495. [CrossRef]
26. Andrade, P.; Silva, I.; Silva, M.; Flores, T.; Cassiano, J.; Costa, D.G. A TinyML Soft-Sensor Approach for Low-Cost Detection and Monitoring of Vehicular Emissions. *Sensors* **2022**, *22*, 3838. [CrossRef] [PubMed]
27. Sanchez-Iborra, R.; Skarmeta, A.F. TinyML-Enabled Frugal Smart Objects: Challenges and Opportunities. *IEEE Circ. Syst. Mag.* **2020**, *20*, 4–18. [CrossRef]
28. Raza, W.; Osman, A.; Ferrini, F.; Natale, F.D. Energy-Efficient Inference on the Edge Exploiting TinyML Capabilities for UAVs. *Drones* **2021**, *5*, 127. [CrossRef]
29. Dutta, D.L.; Bharali, S. TinyML Meets IoT: A Comprehensive Survey. *Internet Things* **2021**, *16*, 100461. [CrossRef]
30. Costa, D.G.; Peixoto, J.P.J.; Jesus, T.C.; Portugal, P.; Vasques, F.; Rangel, E.; Peixoto, M. A Survey of Emergencies Management Systems in Smart Cities. *IEEE Access* **2022**, *10*, 61843–61872. [CrossRef]
31. Blind, K.; Schubert, T. Estimating the GDP effect of Open Source Software and its complementarities with R&D and patents: Evidence and policy implications. *J. Technol. Transf.* **2023**, 1–26. [CrossRef]
32. Fershtman, C.; Gandal, N. Direct and indirect knowledge spillovers: The "social network" of open-source projects. *Rand. J. Econ.* **2011**, *42*, 70–91. [CrossRef]
33. Xiang, X.; Cai, H.; Lam, S.; Pei, Y. International knowledge spillover through co-inventors: An empirical study using Chinese assignees' patent data. *Technol. Forecast. Soc.* **2013**, *80*, 161–174. [CrossRef]

34. Wang, J.; Peng, X. A Study of Patent Open Source Strategies Based on Open Innovation: The Case of Tesla. *Open J. Soc. Sci.* **2020**, *8*, 386–394. [CrossRef]
35. West, J. How open is open enough?: Melding proprietary and open source platform strategies. *Res. Policy* **2003**, *32*, 1259–1285. [CrossRef]
36. West, J.; Bogers, M. Leveraging External Sources of Innovation: A Review of Research on Open Innovation. *J. Prod. Innov. Manag.* **2014**, *31*, 814–831. [CrossRef]
37. Dehghani, M.; Mashatan, A.; Kennedy, R.W. Innovation within networks—Patent strategies for blockchain technology. *J. Bus. Ind. Mark.* **2021**, *36*, 2113–2125. [CrossRef]
38. Theyel, N. Extending open innovation throughout the value chain by small and medium-sized manufacturers. *Int. Small Bus. J.* **2012**, *31*, 256–274. [CrossRef]
39. Gal, M.S. Viral Open Source: Competition vs. Synergy. *J. Compet. Law Econ.* **2012**, *8*, 469–506. [CrossRef]
40. Mies, R.; Bonvoisin, J.; Jochem, R. Harnessing the Synergy Potential of Open Source Hardware Communities. In *Co-Creation: Reshaping Business and Society in the Era of Bottom-Up Economics*; Redlich, T., Moritz, M., Wulfsberg, J.P., Eds.; Springer International Publishing: Cham, Swtizerland, 2019; pp. 129–145.
41. Isaacs, R. *Differential Games: A Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization*; Wiley: New York, NY, USA, 1965.
42. Bellman, R. *Dynamic Programming and Its Application to Variational Problems in Mathematical Economics*; Rand Corporation: Santa Monica, CA, USA, 1958.
43. Jørgensen, S.; Zaccour, G. Developments in differential game theory and numerical methods: Economic and management applications. *Comput. Manag. Sci.* **2007**, *4*, 159–181. [CrossRef]
44. Yang, Y.; Xu, X. A differential game model for closed-loop supply chain participants under carbon emission permits. *Comput. Ind. Eng.* **2019**, *135*, 1077–1090. [CrossRef]
45. Sun, R.; He, D.; Yan, J.; Tao, L. Mechanism Analysis of Applying Blockchain Technology to Forestry Carbon Sink Projects Based on the Differential Game Model. *Sustainability* **2021**, *13*, 11697. [CrossRef]
46. Wei, J.; Wang, C. Improving interaction mechanism of carbon reduction technology innovation between supply chain enterprises and government by means of differential game. *J. Clean. Prod.* **2021**, *296*, 126578. [CrossRef]
47. Yigitcanlar, T. *Technology and the City: Systems, Applications and Implications*; Routledge: London, UK, 2016.
48. Rivas, S.; Urraca, R.; Bertoldi, P.; Thiel, C. Towards the EU Green Deal: Local key factors to achieve ambitious 2030 climate targets. *J. Clean Prod.* **2021**, *320*, 128878. [CrossRef]
49. Skjrseth, J.B. Towards a European Green Deal: The evolution of EU climate and energy policy mixes. *Int. Environ. Agreem.* **2021**, *21*, 25–41. [CrossRef]
50. Chauhan, A.; Jakhar, S.K.; Chauhan, C. The interplay of circular economy with industry 4.0 enabled smart city drivers of healthcare waste disposal. *J. Clean. Prod.* **2020**, *279*, 123854. [CrossRef] [PubMed]
51. D'Amico, G.; L'Abbate, P.; Liao, W.; Yigitcanlar, T.; Ioppolo, G. Understanding Sensor Cities: Insights from Technology Giant Company Driven Smart Urbanism Practices. *Sensors* **2020**, *20*, 4391. [CrossRef]
52. Geels, F.W. Socio-technical transitions to sustainability: A review of criticisms and elaborations of the Multi-Level Perspective. *Curr. Opin. Environ. Sustain.* **2019**, *39*, 187–201. [CrossRef]
53. Bellandi, M.; De Propris, L. Local productive systems' transitions to industry 4.0+. *Sustainability* **2021**, *13*, 13052. [CrossRef]
54. Zhai, X.; An, Y. Analyzing influencing factors of green transformation in China's manufacturing industry under environmental regulation: A structural equation model. *J. Clean. Prod.* **2020**, *251*, 119760. [CrossRef]
55. Acs, Z.; Sanders, M. Endogenous Growth Theory and Regional Extensions. In *Handbook of Regional Science*; Fischer, M.M., Nijkamp, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2021; pp. 615–634.
56. Faisal, A.; Yigitcanlar, T.; Kamruzzaman, M.; Currie, G. Understanding autonomous vehicles: A systematic literature review on capability, impact, planning and policy. *J. Transp. Land Use* **2019**, *12*, 45–72. [CrossRef]
57. Yigitcanlar, T.; Corchado, J.M.; Mehmood, R.; Li, R.Y.M.; Mossberger, K.; Desouza, K. Responsible Urban Innovation with Local Government Artificial Intelligence (AI): A Conceptual Framework and Research Agenda. *J. Open Innov. Technol. Mark. Complex.* **2021**, *7*, 71. [CrossRef]
58. Yigitcanlar, T.; Kankanamge, N.; Regona, M.; Maldonado, A.R.; Rowan, B.; Ryu, A.; Desouza, K.C.; Corchado, J.M.; Mehmood, R.; Li, R.Y.M. Artificial Intelligence Technologies and Related Urban Planning and Development Concepts: How Are They Perceived and Utilized in Australia? *J. Open Innov. Technol. Mark. Complex.* **2020**, *6*, 187. [CrossRef]
59. Dennis, S.; Paz, A.; Yigitcanlar, T. Perceptions and Attitudes Towards the Deployment of Autonomous and Connected Vehicles: Insights from Las Vegas, Nevada. *J. Urban. Technol.* **2021**, *28*, 75–95. [CrossRef]
60. Yigitcanlar, T.; Han, H.; Kamruzzaman, M.; Ioppolo, G.; Sabatini-Marques, J. The making of smart cities: Are Songdo, Masdar, Amsterdam, San Francisco and Brisbane the best we could build? *Land Use Policy* **2019**, *88*, 104187. [CrossRef]
61. Alajlan, N.N.; Ibrahim, D.M. TinyML: Enabling of Inference Deep Learning Models on Ultra-Low-Power IoT Edge Devices for AI Applications. *Micromachines* **2022**, *13*, 851. [CrossRef]
62. Garcia-Retuerta, D.; Chamoso, P.; Hernández, G.; Guzmán, A.S.R.; Yigitcanlar, T.; Corchado, J.M. An Efficient Management Platform for Developing Smart Cities: Solution for Real-Time and Future Crowd Detection. *Electronics* **2021**, *10*, 765. [CrossRef]

63. Ndiaye, M.; Oyewobi, S.S.; Abu-Mahfouz, A.M.; Hancke, G.P.; Kurien, A.M.; Djouani, K. IoT in the Wake of COVID-19: A Survey on Contributions, Challenges and Evolution. *IEEE Access* **2020**, *8*, 186821–186839. [CrossRef]
64. Prakash, S.; Stewart, M.; Banbury, C.; Mazumder, M.; Warden, P.; Plancher, B.; Reddi, V.J. Is TinyML Sustainable? Assessing the Environmental Impacts of Machine Learning on Microcontrollers. *arXiv* **2023**, arXiv:2301.11899.
65. Das, K.P.; Chandra, J. A survey on artificial intelligence for reducing the climate footprint in healthcare. *Energy Nexus* **2023**, *9*, 100167. [CrossRef]
66. Istofa; Prajitno, P.; Susila, I.P.A. Systematic Literature Review of TinyML for Environmental Radiation Monitoring System. In Proceedings of the 6th Mechanical Engineering, Science and Technology International Conference (MEST 2022), Okinawa, Japan, 20–21 December 2022; Atlantis Press: Amsterdam, The Netherlands, 2023; pp. 461–473.
67. Antonini, M.; Pincheira, M.; Vecchio, M.; Antonelli, F. An Adaptable and Unsupervised TinyML Anomaly Detection System for Extreme Industrial Environments. *Sensors* **2023**, *23*, 2344. [CrossRef] [PubMed]
68. Vuppalapati, C.; Ilapakurti, A.; Chillara, K.; Kedari, S.; Mamidi, V. Automating Tiny ML Intelligent Sensors DevOPS Using Microsoft Azure. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 10–13 December; pp. 2375–2384.
69. Zhang, J.; Chen, C.; Peng, J.; Liang, J.; Nikhath, K. Early warning method and system of building environmental security based on TinyML and CloudML technology. In Proceedings of the 2nd International Conference on Information Technology and Intelligent Control (CITIC 2022), Kunming, China, 15–17 July 2022.
70. Rathnappriya, R.H.K.; Sakai, K.; Okamoto, K.; Kimura, S.; Haraguchi, T.; Nakandakari, T.; Setouchi, H.; Bandara, W.B.M.A.C. Global Sensitivity Analysis of Key Parameters in the APSIMX-Sugarcane Model to Evaluate Nitrate Balance via Treed Gaussian Process. *Agronomy* **2022**, *12*, 1979. [CrossRef]
71. Helman, D.; Yungstein, Y.; Mulero, G.; Michael, Y. High-Throughput Remote Sensing of Vertical Green Living Walls (VGWs) in Workplaces. *Remote Sens.* **2022**, *14*, 3485. [CrossRef]
72. Balaska, V.; Adamidou, Z.; Vryzas, Z.; Gasteratos, A. Sustainable Crop Protection via Robotics and Artificial Intelligence Solutions. *Machines* **2023**, *11*, 774. [CrossRef]
73. Sishodia, R.P.; Ray, R.L.; Singh, S.K. Applications of Remote Sensing in Precision Agriculture: A Review. *Remote Sens.* **2020**, *12*, 3136. [CrossRef]
74. Ihoume, I.; Tadili, R.; Arbaoui, N.; Benchrifa, M.; Idrissi, A.; Daoudi, M. Developing a TinyML-Oriented Deep Learning Model for an Intelligent Greenhouse Microclimate Control from Multivariate Sensed Data. In *Intelligent Sustainable Systems, Singapore*; Nagar, A.K., Singh Jat, D., Mishra, D.K., Joshi, A., Eds.; Springer Nature: Singapore, 2023; pp. 283–291.
75. Macedo, F.L.; Nóbrega, H.; de Freitas, J.G.; Ragonezi, C.; Pinto, L.; Rosa, J.; Pinheiro de Carvalho, M.A. Estimation of Productivity and Above-Ground Biomass for Corn (*Zea mays*) via Vegetation Indices in Madeira Island. *Agriculture* **2023**, *13*, 1115. [CrossRef]
76. Neethirajan, S. SOLARIA-SensOr-driven resiLient and adaptive monitoRIng of farm Animals. *Agriculture* **2023**, *13*, 436. [CrossRef]
77. Muthuramalingam, S.; Bharathi, A.; Rakesh Kumar, S.; Gayathri, N.; Sathiyaraj, R.; Balamurugan, B. *IoT Based Intelligent Transportation System (IoT-ITS) for Global Perspective: A Case Study*; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 279–300.
78. Eshghi, H.; Zamen, M.; Kahani, M. Energy and environmental investigation on photovoltaic system performance by application of square cross-sectional two-phase closed thermosyphon. *Environ. Sci. Pollut. Res.* **2023**. [CrossRef] [PubMed]
79. Ren, H.; Anicic, D.; Runkler, T.A. TinyReptile: TinyML with Federated Meta-Learning. In Proceedings of the The International Joint Conference on Neural Network (IJCNN) 2023, Gold Coast, Australia, 18–23 June 2023.
80. Osman, A.; Abid, U.; Gemma, L.; Perotto, M.; Brunelli, D. TinyML Platforms Benchmarking. In Proceedings of the International Conference on Applications in Electronics Pervading Industry, Environment and Society, Genova, Italy, 26–27 September 2022; Saponara, S., De Gloria, A., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2022; pp. 139–148.
81. Signoretti, G.; Silva, M.; Andrade, P.; Silva, I.; Sisinni, E.; Ferrari, P. An Evolving TinyML Compression Algorithm for IoT Environments Based on Data Eccentricity. *Sensors* **2021**, *21*, 4153. [CrossRef] [PubMed]
82. Li, T.; Ma, L.; Liu, Z.; Liang, K. Economic Granularity Interval in Decision Tree Algorithm Standardization from an Open Innovation Perspective: Towards a Platform for Sustainable Matching. *J. Open Innov. Technol. Mark. Complex.* **2020**, *6*, 149. [CrossRef]
83. Economides, N.; Katsamakas, E. Two-Sided Competition of Proprietary vs. Open Source Technology Platforms and the Implications for the Software Industry. *Manag. Sci.* **2006**, *52*, 1057–1071. [CrossRef]
84. Maier, A.; Yang, S.H.; Maleki, F.; Muthukrishnan, N.; Forghani, R. Offer Proprietary Algorithms Still Protection of Intellectual Property in the Age of Machine Learning? In Proceedings of the Bildverarbeitung für die Medizin 2022, Wiesbaden, Germany, 26–28 June 2022; Maier-Hein, K., Deserno, T.M., Handels, H., Maier, A., Palm, C., Tolxdorff, T., Eds.; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2022; pp. 345–350.
85. Kamath, V.; Morgan, J.; Ali, M.I. Industrial IoT and Digital Twins for a Smart Factory: An open source toolkit for application design and benchmarking. In Proceedings of the 2020 Global Internet of Things Summit (GIoTS), Dublin, Ireland, 3 June 2020; pp. 1–6.

86. Llanes, G.; de Elejalde, R. Industry equilibrium with open-source and proprietary firms. *Int. J. Ind. Organ.* **2013**, *31*, 36–49. [CrossRef]
87. Zhu, K.X.; Zhou, Z.Z. Research Note-Lock-In Strategy in Software Competition: Open-Source Software vs. Proprietary Software. *Inform. Syst. Res.* **2012**, *23*, 536–545. [CrossRef]
88. August, T.; Chen, W.; Zhu, K. Competition among Proprietary and Open-Source Software Firms: The Role of Licensing in Strategic Contribution. *Manag. Sci.* **2020**, *67*, 3041–3066. [CrossRef]
89. Kilamo, T.; Hammouda, I.; Mikkonen, T.; Aaltonen, T. From proprietary to open source—Growing an open source ecosystem. *J. Syst. Softw.* **2012**, *85*, 1467–1478. [CrossRef]
90. Boulanger, A. Open-source versus proprietary software: Is one more reliable and secure than the other? *IBM Syst. J.* **2005**, *44*, 239–248. [CrossRef]
91. Lindenschmidt, K.-E. RIVICE—A Non-Proprietary, Open-Source, One-Dimensional River-Ice Model. *Water* **2017**, *9*, 314. [CrossRef]
92. O'Mahony, S.; Karp, R. From proprietary to collective governance: How do platform participation strategies evolve? *Strateg. Manag. J.* **2022**, *43*, 530–562. [CrossRef]
93. Rubinstein, A. Comments on the Interpretation of Game Theory. *Econometrica* **1991**, *59*, 909–924. [CrossRef]
94. Niculescu, M.F.; Wu, D.J.; Xu, L. Strategic Intellectual Property Sharing: Competition on an Open Technology Platform under Network Effects. *Inf. Syst. Res.* **2018**, *29*, 498–519. [CrossRef]
95. Rasmusen, E. *Games and Information: An Introduction to Game Theory*, 4th ed.; Wiley-Blackwell: Hoboken, NJ, USA, 2006; p. 560.
96. Aurell, A.; Carmona, R.; Laurière, M. Stochastic Graphon Games: II. The Linear-Quadratic Case. *Appl. Math. Optim.* **2022**, *85*, 39. [CrossRef]
97. Fu, X. How does openness affect the importance of incentives for innovation? *Res. Policy* **2012**, *41*, 512–523. [CrossRef]
98. Lee, R.; Lee, J.; Garrett, T.C. Synergy effects of innovation on firm performance. *J. Bus. Res.* **2019**, *99*, 507–515. [CrossRef]
99. Haddad, W.M.; Nersesov, S.G.; Du, L. Finite-time stability for time-varying nonlinear dynamical systems. In Proceedings of the 2008 American Control Conference, Seattle, WA, USA, 11–13 June 2008; pp. 4135–4139.
100. Meshkov, V.Z. On the possible rate of decay at infinity of solutions of second order partial differential equations. *Math. USSR-Sb.* **1992**, *72*, 343–361. [CrossRef]
101. Mcmillan, G.S.; Mauri, A.; Casey, D.L. The scientific openness decision model: "Gaming" the technological and scientific outcomes. *Technol. Forecast. Soc.* **2014**, *86*, 132–142. [CrossRef]
102. Chen, J.; Wei, Z.; Liu, J.; Zheng, X. Technology sharing and competitiveness in a Stackelberg model. *J. Compet.* **2021**, *13*, 5–20. [CrossRef]
103. Boyle, J. Open Source Innovation, Patent Injunctions, and the Public Interest. *Duke Law Technol. Rev.* **2012**, *11*, 30. [CrossRef]
104. Rayna, T.; Striukova, L. Large-scale open innovation: Open source vs. patent pools. *Int. J. Technol. Manag.* **2010**, *52*, 477–496. [CrossRef]
105. Adner, R.; Chen, J.; Zhu, F. Frenemies in Platform Markets: Heterogeneous Profit Foci as Drivers of Compatibility Decisions. *Manag. Sci.* **2020**, *66*, 2432–2451. [CrossRef]
106. Carrasco-Farre, C.; Snihur, Y.; Berrone, P.; Enric Ricart, J. The stakeholder value proposition of digital platforms in an urban ecosystem. *Res. Policy A J. Devoted Res. Policy Res. Manag. Plan.* **2022**, *51*, 104488. [CrossRef]
107. Chari, M.; Steensma, H.K.; Connaughton, C.; Heidl, R. The influence of patent assertion entities on inventor behavior. *Strateg. Manag. J.* **2022**, *43*, 1666–1690. [CrossRef]
108. Ruokolainen, J.; Nätti, S.; Juutinen, M.; Puustinen, J.; Holm, A.; Vehkaoja, A.; Nieminen, H. Digital healthcare platform ecosystem design: A case study of an ecosystem for Parkinson's disease patients. *Technovation* **2023**, *120*, 102551. [CrossRef]
109. Shree, D.; Kumar Singh, R.; Paul, J.; Hao, A.; Xu, S. Digital platforms for business-to-business markets: A systematic review and future research agenda. *J. Bus. Res.* **2021**, *137*, 354–365. [CrossRef]
110. Kretschmer, T.; Leiponen, A.; Schilling, M.; Vasudeva, G. Platform ecosystems as meta-organizations: Implications for platform strategies. *Strateg. Manag. J.* **2022**, *43*, 405–424. [CrossRef]
111. Holgersson, M.; Wallin, M.W. The patent management trichotomy: Patenting, publishing, and secrecy. *Manag. Decis.* **2017**, *55*, 1087–1099. [CrossRef]
112. Remneland Wikhamn, B.; Styhre, A. Open innovation ecosystem organizing from a process view: A longitudinal study in the making of an innovation hub. *RD Manag.* **2023**, *53*, 24–42. [CrossRef]
113. Zhao, X.; Guo, Y.; Mi, J. Value creation and value capture from patents: Theory and implications for patent strategies. *J. Innov. Knowl.* **2023**, *8*, 100397. [CrossRef]
114. Holgersson, M.; Granstrand, O. Value capture in open innovation markets: The role of patent rights for innovation appropriation. *Eur. J. Innov. Manag.* **2022**, *25*, 320–339. [CrossRef]
115. Gambardella, A. Private and social functions of patents: Innovation, markets, and new firms. *Res. Policy* **2023**, *52*, 104806. [CrossRef]

116. Alam, M.A.; Rooney, D.; Taylor, M. From ego-systems to open innovation ecosystems: A process model of inter-firm openness. *J. Prod. Innov. Manag.* **2022**, *39*, 177–201. [CrossRef]
117. Cavallo, A.; Burgers, H.; Ghezzi, A.; van de Vrande, V. The evolving nature of open innovation governance: A study of a digital platform development in collaboration with a big science centre. *Technovation* **2022**, *116*, 102370. [CrossRef]

sustainability

Article

Does the Digital Economy Promote Coordinated Urban–Rural Development? Evidence from China

Yue Cheng * and Dong Zheng

School of Public Administration, Guangxi University, Nanning 530004, China
* Correspondence: 20110003@gxu.edu.cn

Abstract: Urban–rural coordination development is a key factor in achieving sustainable development. The research sample consisted of panel data for 30 provinces in China for the period from 2011 to 2020. Our aim was to investigate whether and how the digital economy affects coordinated urban–rural development by using a panel data model, a spatial Durbin model (SDM), and a mediating effects model. The results indicate that (1) the growth of the digital economy has increased the level of coordinated urban–rural development directly and indirectly; (2) the coordinated development of urban and rural areas and the spatial distribution of the digital economy are highly correlated, with eastern regions generally experiencing a high level of agglomeration and central and western regions having a low level of agglomeration; (3) the digital economy can promote coordinated urban–rural development by reducing the income gap between urban and rural areas; and (4) the direct and the spatial promotion effects of digital economy development on coordinated urban–rural development appear to be stronger in the eastern region, insignificant in the central region, and to have a significant direct inhibition, as well as a significant spatial spillover effect, in the western region. This study provides a reference for China and other developing countries similar to China on how to promote coordinated urban and rural development in the development process of the digital economy.

Keywords: digital economy; urban–rural dualism; coordinated urban–rural development; urban–rural income gap; spatial spillover effects

Citation: Cheng, Y.; Zheng, D. Does the Digital Economy Promote Coordinated Urban–Rural Development? Evidence from China. *Sustainability* **2023**, *15*, 5460. https://doi.org/10.3390/su15065460

Academic Editor: Antonella Petrillo

Received: 27 February 2023
Revised: 17 March 2023
Accepted: 17 March 2023
Published: 20 March 2023

1. Introduction

China's urban–rural structure has undergone tremendous changes since the mid-1980s, and since 2003, the government has implemented sustainability policies for coordinated urban–rural development [1]. China has now achieved its goal of poverty alleviation, and the inequity between urban and rural development has been dramatically reduced. However, the imbalance between urban and rural development still exists in some remote regions [2]. Urban-biased policies and the urban–rural dual system are the primary causes of the urban–rural gap [3]. There are three main representative theories on the development of urban–rural relationships [4]: the urban–rural connection theory is represented by the urban–rural integration of utopian socialism and Marxism, the Lewis–Ranis–Fei model represents the urban–rural dual structure, and the Desakota model and the regional network model represent the urban–rural coordinated development [5,6]. Fostering urban–rural interdependence is seen as an effort to support sustainable urban–rural and regional growth [7,8].

With the development of next-generation technologies such as mobile internet, cloud computing, big data, the Internet of Things, blockchain, and artificial intelligence, China's economy is driven toward high-quality development by the broad and rapidly expanding digital economy. According to the "White Paper on China's Digital Economy Growth", published by the China Academy of Information and Communication Technology (CAICT) in 2022, the Chinese government is committed to fostering the expansion of its digital economy. Since 2012, the Chinese digital economy's average annual growth rate has been 15.9%, significantly higher than the average annual growth rate of China's GDP over the same period. In 2021, the digital economy reached CNY 45.5 trillion, representing a nominal gain of 16.2% annually. The widespread use of digital technologies has triggered an economic revolution and brought new ways of practising production, and the digitisation of economic systems is becoming increasingly important. From the point of view of technological progress, digitalisation causes economic activities to have increasing marginal returns, breaking the law of decreasing returns for each additional unit of a factor input after the input of that factor reaches a critical point in the industrial economy. From the perspective of production organisation, digitalisation can significantly reduce transaction costs. The transparency of the network and the openness of information in the digital era have greatly reduced the marginal costs of market transactions; boundaries of enterprises are shrinking, transactions and cooperation between enterprises are becoming more frequent, and flat production organisation forms have emerged, reducing the cost burden of enterprises. From the perspective of resource allocation, in the digital economy, the problem of market failure is alleviated to a certain extent, and the role of market regulation is enhanced. From the perspective of the division of labour, the antagonism between urban and rural relations is diminishing. With the proliferation of information and communication technologies, the high-value-added segments of the industrial chain, such as research, development, and sales, are gradually moving closer to technology-intensive cities, while the low-value-added segments, such as production and processing, are moving to labour-intensive townships. In this process, cities and townships brought into play their comparative advantages and deepened their collaborative relationship, changing the dichotomy between the urban economy of industrial production and the agricultural economy of smallholder production and forming a new pattern of mutually beneficial and complementary urban–rural division of labour [9]. The growth of the digital economy will have a profound impact on reshaping the new urban–rural relationship, achieving balanced development in urban and rural areas and changing the pattern of income distribution between urban and rural areas [10]. Therefore, the attention of numerous scholars has been drawn to how to effectively promote coordinated urban–rural development with digital economic growth. Most scholars study the impact of the digital economy on coordinated urban–rural development from the perspective of the gap between urban–rural income and consumption. Some scholars have noted the importance of digitisation in the public sector and that digital public platforms can provide better and equal access to public services across different sectors, which can reduce divisions and inequalities between countries, the private and public sectors, and urban and rural areas [11,12]. However, the academic community has yet to determine whether the expansion of the digital economy would enable the "digital dividend" and thus promote coordinated urban–rural development or whether it would worsen the "digital divide" and, in that way, inhibit coordinated urban–rural development [13]; their findings are still highly controversial.

The following are possible contributions of this study: (1) the impact of the digital economy on the coordinated development of urban and rural areas and its mechanism of action are explored from the perspective of narrowing the urban–rural gap in the context of the rural revitalisation strategy; (2) China's innovative evaluation index system of digital economy level is constructed from four dimensions: digital economy infrastructure (DIS) support, digital economy innovation and entrepreneurship (DIE) level, digital talent pool (DTP), and digital technology services (DTS); and (3) the impact of the digital economy on coordinated urban–rural development is examined from the perspective of the spatial spillover effect, and this examination also combines the direct and the spatial heterogeneity to further improve and complement the existing research.

The study is arranged as follows. Section 2 is a literature review, Section 3 introduces the logical mechanism and research hypotheses, Section 4 describes the data and methods, Section 5 provides the empirical results, Section 6 provides conclusions, and Section 7 provides policy recommendations and limitations. The research framework is shown in Figure 1.

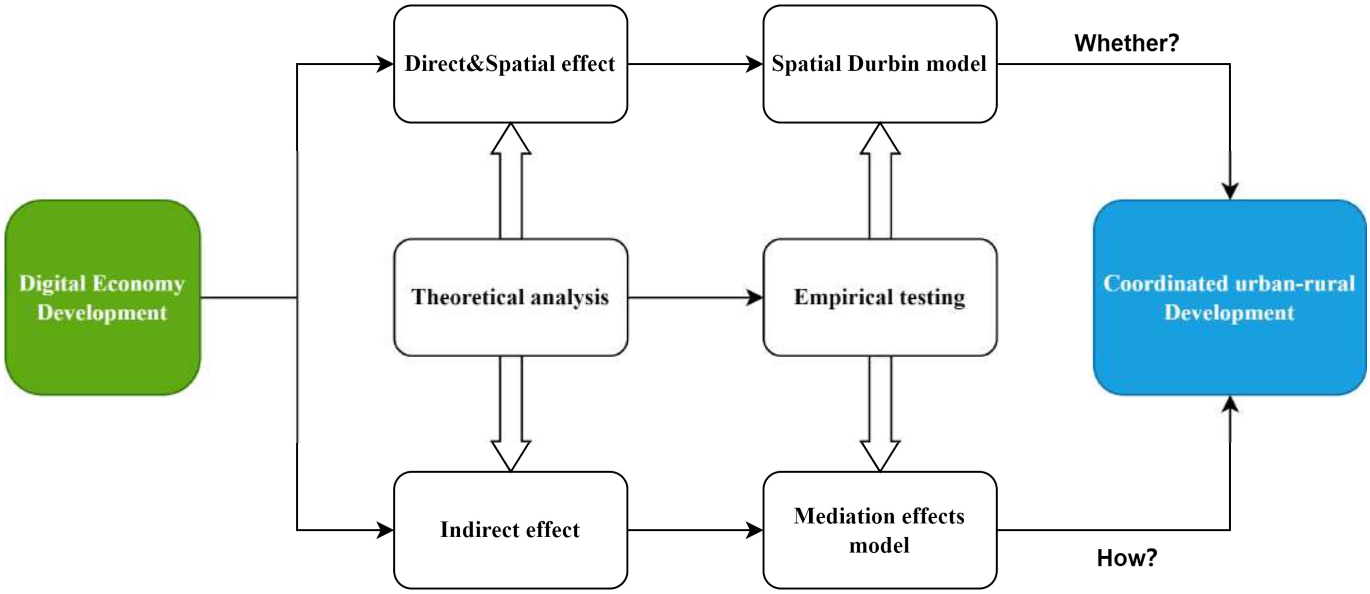


Figure 1. Research framework.

2. Literature Review

China's fast-rising digital economy has recently emerged as the "new engine" of the economic and social revolution. This has caused the research in this area to exponentially expand. Most of the relevant literature on the digital economy has observed it from three different perspectives—theory, mechanism, and realisation. This scope of literature has brought us a more extensive understanding of how the digital economy affects high-quality development [14,15], a circular and sustainable economy [16,17], green innovation [18,19], the transformation and upgrading of industrial structure [20,21], and total factor productivity [22]. Digitalisation brings opportunities as well as challenges, and digital technologies have contributed to a shift in household financial models and have required financial institutions to accelerate the pace of innovation to adapt to the changing environment [23]. Digitalisation has enhanced the international competitiveness of businesses and has had a positive impact on the economies of countries at all levels of development [24,25]. However, the digitalisation of the economy has also triggered intense market competition and unfair practices [26], which require governments to adopt scientific policies to address these issues. At present, there are primarily three distinct viewpoints when attempting to precisely observe the digital economy's influence on the coordinated growth of urban and rural areas.

From the first viewpoint, the sharing aspect of the digital economy can support the sensible allocation of resources between urban and rural areas, narrowing the income gap between urban and rural inhabitants and promoting coordinated urban–rural development [27]. This viewpoint is supported by the fact that the digital economy directly decreases the urban–rural gap through the impact it has on market integration [28], as well as through the modular division of the labour effect. What is more, the agglomeration economy indirectly reduces the urban–rural gap via workforce reallocation and the agglomeration effect [29]. What has also been stated in the context of this viewpoint is that even though the digital economy has surpassed its original time and space limitations, it has still yielded the expansion of employment opportunities [30]. The digital economy's spillover effects have generated a significant number of jobs suited for the skill levels of farmers while parallelly raising their incomes, and, thereby, enhancing the market resource allocation efficiency. Based on their empirical research, Zhou (2022) came to the conclusion that [31], with the reform of the household registration system and the construction of transport infrastructure, the two-way flow of the urban and the rural factors can extend the optimal allocation effect that the digital economy has on urban and rural incomes and further promote the development of the digital economy itself.

From the second viewpoint, digital technology will restrict coordinated urban and rural development. The "digital divide" between the urban and rural areas can nowadays be characterised by the vast difference in their digital infrastructure and their populations' digital literacy. On the one side, the digital industry is more concentrated in the urban areas where economic activities normally take place due to the digital infrastructure's higher quality and its higher level of advancement. On the other side, the average education level of the rural inhabitants falls behind that of the residents of the urban areas [32,33]. As an additional point, digital literacy, digital information absorption, and digital knowledge digestion skills are not particularly strong among rural inhabitants either. It should be noted as well that Jun (2017) found that digitisation and the information revolution have not lessened the gap between the rich and the poor as was anticipated [34], but they have rather resulted in the widening of the urban–rural income gap, recognised via the Matthew effect. Based on the empirical tests that they have conducted, Yaping (2019) found that [35], although the Internet's high efficiency has reduced the cost of searching and acquiring information, and even though it has increased income levels, due to the disparity in the farmers' levels of Internet application, the reduced cost of searching the Internet is not significant in the rural areas, and this further widens the income gap between the urban and rural regions.

According to the third viewpoint, the effect of the digital economy on the urban–rural development gap follows an inverted U-shaped pattern [36]. In other words, the digital economy has altered the traditional labour market's growth pattern, and it has further optimised the structure of income distribution. China's digital economy is still undergoing rapid development, while some simple and mechanised jobs have disappeared because of digital technologies such as artificial intelligence and many low-skilled jobs have been created, giving low-skilled and middle-skilled workers more employment opportunities and allowing rural labourers to earn higher wages. This has in turn reduced the urban–rural wage gap and further decreased the income disparity between them. Looking from a long-term standpoint, however, further development of the digital economy can lead to the opposite result in the future [37]. More specifically, the level of knowledge and the technical skills that will be required in the future will increase together with digital improvement, which will then leave the low-skilled labourers to face the double risk of losing employment opportunities due to possibly being substituted by artificial intelligence or their insufficient levels of digital literacy. Subsequently, this leads to a reduction in employment options for low-skilled rural labourers and the majority of the farmers who do not meet the job skill requirements and who will once again find themselves unemployed [38].

In summary, the existing literature has deeply studied the relationship between the digital economy and coordinated urban–rural development, thus providing a solid foundation for our study. However, there are still shortcomings in terms of the research content and perspective. First, most of the existing research focuses on the definition and measurement of the digital economy or coordinated urban–rural development, while studies of the combination of these two concepts are lacking. Second, research on the impact of the digital economy on coordinated urban–rural development and its mechanisms has yet to be established and improved. Third, existing research has only examined the regional heterogeneity of the direct effects of the digital economy on coordinated urban–rural development, ignoring the regional heterogeneity of the spatial effects of the digital economy on coordinated urban–rural development.

To fill the gaps in current studies, we aim to combine the digital economy and coordinated urban–rural development and investigate the influence of the relationship between them, with the objective of providing empirical support for one of the three different conclusions mentioned above. Additionally, it is hoped that our research from a spatial perspective will lead to a different conclusion from that obtained in existing studies. Therefore, this study uses panel data of 30 Chinese provinces from 2011 to 2020 to systematically explore the spatial impact, action mechanism, and heterogeneity of the digital economy impacts on coordinated urban–rural development. This is achieved by constructing a spatial Durbin model (SDM) and a mediating effects model and by proposing scientific and targeted policy recommendations.

3. Theoretical Analysis and Research Hypothesis

3.1. The Direct Effect of the Digital Economy on the Coordination of the Urban and Rural Development

The mechanism by which the digital economy facilitates coordinated urban–rural development manifests itself in the three following ways.

Firstly, the digital economy significantly improves the farmers' ability to collect and access information, lowering their cost of learning and knowledge sharing and thus contributing to the optimisation and the upgrading of rural industries. Put differently, the digital economy, with its technology, has significantly decreased the economic and social transaction costs for businesses, individuals, and the public sector by reducing the cost of information search [39].

Secondly, the digital economy allows rural areas to have equal access to an increasing quantity of high-quality public service resources. Digital technology is a catalyst not only for economic transformation but also for social transformation as well. It provides significant social benefits, particularly in terms of facilitating access to basic services, such as financial services and education. Furthermore, the digital economy addresses the lack of traditional service provision in underdeveloped rural areas, and it fosters the coordination and rapid improvement of the public service levels in both urban and rural areas [40].

Thirdly, the digital economy can increase the rural population's consumption capacity and income. The knowledge-sharing characteristics of the digital economy allow it to maximize resource allocation and efficiency optimisation and offer consumers a greater product choice and cheaper access to goods of identical quality [41]. When it comes to the role of the digital economy in agricultural development, digital agriculture can effectively reduce the wealth gap between urban and rural residents and, at the same time, increase the farmers' disposable income, thereby contributing to the improvement of their overall social welfare level [42]. Taking all of this into consideration, the following hypothesis is proposed:

H1. *The growth of the digital economy has had an impact on the increase in the degree of urban and rural development coordination.*

3.2. Digital Economy's Spatial Spillover Effects on Coordinated Urban–Rural Development

According to the first law of geography, geographical objects or attributes are interrelated and dependent on one another in terms of their spatial distribution. Analogously, there is agglomeration, random and regular distribution [43], and spatial spillover and spatial dependence within the digital economy sphere as well. The primary manifestations of the digital economy's spillover effects are knowledge spillover and human capital spillover. When it comes to knowledge spillover, rural areas have surpassed the traditional methods of acquiring and learning information through education and technical training. Instead, they now construct digital villages. Using digital technologies, such as the Internet and augmented virtual reality, farmers can acquire and gain knowledge more quickly and easily now, and, in that way, they are allowing for a gradual shift in the mentality and cognitive level of the rural communities.

When it comes to human capital spillover, the development of the digital economy has resulted in the creation of a substantial number of job opportunities. Farmers have improved their professional level and their abilities due to the knowledge spillover effect. They also have the opportunity, through the Internet, to obtain a great amount of information concerning the jobs that match their abilities, which has in turn increased the efficiency of the resource allocation market and facilitated the flow of human capital.

The second manifestation of the spatial dependence of the digital economy's development is the imperfect construction of new digital infrastructure and the spatial disparities in its distribution [44]. By constructing a virtual space, governments and businesses, as well as cities and rural areas, can engage in cross-regional cooperation, production, and operation with the costs of coordination and management being reduced. Consequently, this leads to resource sharing and complementary advantages being established between regions and between cities and rural areas [45]. Previous research has indicated that the impact of the digital economy on coordinated regional development has a spatial effect [46]. In other words, the development of the digital economy in an area can stimulate the region's coordinated regional development and the regional development of the province's adjacent regions. What is more, the digital economy has significant spatial spillover effects on the other elements included in the achievement of coordinated urban–rural development, such as urban–rural economic integration and rural revitalisation [47,48].

In addition, due to China's vast size, significant differences in economic resources and other factors have been found to exist between different regions. The digital economy and the urban and rural development in each region may exhibit distinctive characteristics, which then may result in disparate effects of the digital economy on coordinated urban and rural development in different regions.

Considering the preceding analysis, the following research hypothesis is proposed:

H2. *Through the effect of the spatial spillover, the digital economy can boost the level of coordinated urban–rural development in neighbouring regions.*

H3. *The impact of the digital economy on coordinated regional development is regionally heterogeneous.*

3.3. Mediating Mechanisms for the Effects of the Digital Economy on Coordinated Urban and Rural Development

The growth of the digital economy has reduced the income gap between urban and rural residents primarily through the following mechanisms. First, digital technology has empowered industries, which is conducive to adjusting the structure of the agricultural industry, extending the industrial, value, and income chains, and thus increasing the income of rural residents [49]. Second, there is a gap in Internet penetration, digital technology application, and e-commerce development between urban and rural areas. Most rural areas can use the "latecomer advantage" to fully implement the digital economy dividend and to close the income gap with urban residents [50]. Thirdly, the application of the Internet and big data in agriculture can reduce agricultural production costs, increase access to information, and promote the coordinated development of the entire industrial chain (production, processing, and marketing of agricultural products). It can also lead to the development of services related to agriculture such as recreation and tourism, which can increase the income levels of rural residents [51]. Fourthly, the development of the digital economy can create numerous employment opportunities in rural areas, thereby expanding the employment base and enhancing the employment quality and income levels of locals [52].

H4. *Development of the digital economy promotes coordinated urban–rural development by narrowing the income gap between urban and rural areas.*

4. Methodology and Design

4.1. Methodology

This study uses a panel data model, a spatial econometric model, and a mediating effects model to investigate the impact and mechanisms of the digital economy on the detection of urban–rural coordination.

Firstly, a panel data model is used to verify whether the digital economy has an impact on coordinated urban–rural development and whether the effect is positive or negative, thereby providing a basis for subsequent spatial econometric analysis.

Second, a spatial econometric model is used to verify whether there is a spatial spillover effect of the digital economy on urban–rural coordination and whether the effect is positive or negative, as well as to further explore the regional heterogeneity of direct and spatial effects.

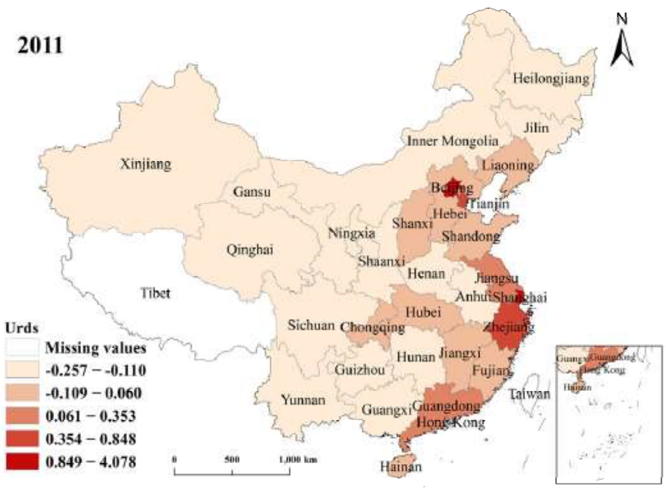
Third, a mediating effects model is used to verify whether the digital economy can promote coordinated urban–rural development by reducing the urban–rural income gap.

Finally, robustness analysis was conducted using three methods, i.e., 1% tail-shrinking on the core explanatory variables, replacement of the core variables, and replacement of the spatial matrix, in order to ensure the reliability and stability of the study results.

4.2. Variable Selection and Description

4.2.1. Measuring the Level of the Coordinated Urban and Rural Development

The coordinated urban–rural development's spatial distribution by province in China in 2011 and 2020 is depicted in Figure 2. Currently, the Gini coefficient (Gini) and the urban–rural binary contrast index (Duci) are seen as the most important indicators of the coordinated development of urban and rural areas. The Gini coefficient is applicable to the evaluation of the overall income gap, but it is, at the same time, insensitive to the income structure differences between urban and rural areas. The urban–rural dichotomy index is used in explaining and analysing the dichotomous economic structure from the perspective of the economic development process of transforming an agricultural economy into a modern industrial economy. It is more suitable for measuring the degree of coordinated urban–rural development.



(**a**) (**b**)

Figure 2. Levels of coordinated urban–rural development in Chinese provinces in 2011 and 2020. (**a**) is the level of coordinated urban–rural development in Chinese provinces in 2011, (**b**) is the level of coordinated urban–rural development in Chinese provinces in 2020.

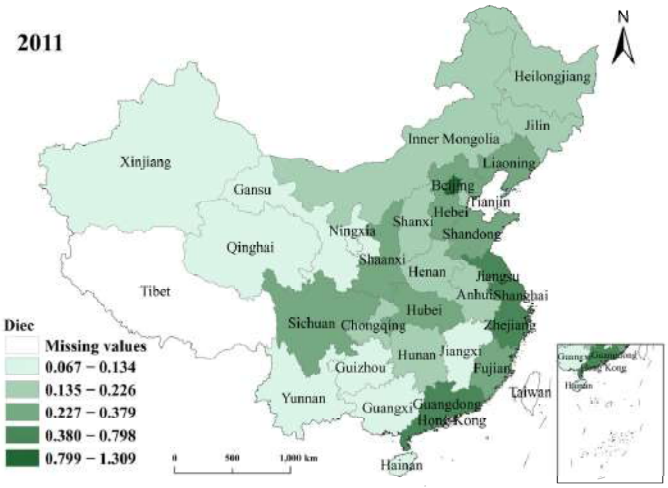
In this paper, we integrate the urban–rural dichotomy contrast index and the proportion of one output value into the evaluation index system of the coordinated urban–rural development level. We also calculate the current final level of coordinated urban–rural development using principal component analysis (PCA). The urban–rural dichotomy contrast index is calculated as shown in the following Equation (1):

$$\mathrm{Duci} = |\mathrm{G1}/\mathrm{G} - \mathrm{L1}/\mathrm{L}| \quad (1)$$

where Duci stands for the rural–urban dichotomy index, G represents the gross regional production, and G1 represents the non-agricultural sector output (the secondary and the tertiary sectors). L stands for total employment, and L1 stands for non-agricultural sector employment.

4.2.2. Measuring the Level of Development of the Digital Economy

There is still no universal agreement on how to measure and evaluate the development level of the digital economy. Scholars primarily evaluate the state of the digital economy in terms of Internet development and digital infrastructure and applications [18,53], failing to consider the importance of digital talent and innovation in the development of the digital economy. In this paper, we develop a regional digital economy measurement index system for China based on four dimensions: (1) digital economy infrastructure support (DIS), (2) level of digital economy innovation and entrepreneurship (DIE), (3) the digital talent pool (DTP), and (4) the digital technology services (DTS). Included are the length of optical fibre cables, the Internet penetration rate, the mobile phone penetration rate, the number of Internet broadband interfaces, the number of Internet domain names, information transmission, computer services, fixed asset investment in the software industry, the number of new enterprises, the attraction of inward investment and venture capital, the number of patents, and the number of patent applications. Using the entropy method, we determined the level of the digital economy. The spatial distribution of digital economy levels by province in China in 2011 and 2020 is shown in Figure 3.



(**a**) (**b**)

Figure 3. Levels of China's provincial digital economy in 2011 and 2020. (**a**) is the level of China's provincial digital economy in 2011, (**b**) is the level of China's provincial digital economy in 2020.

4.2.3. Measuring the Urban–Rural Income Gap

This study uses the Thiel index to measure the urban–rural income gap. The Theil index takes population changes into account, and it is more sensitive to the income changes in both the high- and the low-income groups positioned at the two ends of the dispersion.

4.2.4. Selection of the Control Variables

Based on the selections of the control variables given in the literature [54–57], and to ensure the reliability of the measurement results, we controlled four variables. The first one was the people's livelihood fiscal expenditure, expressed as the proportion of the expenditure on education, health care, housing, social security, and employment in the fiscal budget. The second one represented the years of education per capita, expressed as the average sum of the years of education of the educated population regional groups, calculated via the method shown in Equation (2). The third one was the level of financial development, expressed as the ratio of total deposits and loans to GDP. The fourth and final control variable was the fiscal expenditure on science and technology, expressed as the proportion of GDP in fiscal science and technology expenditures. Table 1 displays the names and the abbreviations of the primary variables.

$$\text{Avsy} = \frac{\begin{array}{c}\text{number of elementary schools} \times 6 + \text{number of junior high schools} \times 9 + \\ \text{number of senior high schools} \times 12 + \text{secondary schools} \times 12 + \text{specialists} \\ \times 15 + \text{bachelor's degrees} \times 16 + \text{graduate students} \times 19\end{array}}{\text{total population over 6 years}} \tag{2}$$

Table 1. Variable selection and description.

Variable Type	Variable Name	Variable Symbol
Explained variables	Level of coordinated urban and rural development	Urds
Core explanatory variables	Digital economy level	Diec
Mediator variables	Theil index	Urig
Control variables	Financial expenditure on people's livelihood Years of education per capita Level of financial development Fiscal expenditure on science and technology	Fepl Avsy Finance Scte

4.3. Data Sources and Descriptive Statistics

Using panel data from 30 provinces (municipalities directly under the Central Government and autonomous regions) from 2011 to 2020, this paper empirically examines the impact of China's digital economy on the coordinated growth of urban and rural areas. Hong Kong, Macao, Taiwan, and Tibet were omitted from the analysis due to insufficient and excessively missing data for some regions in those areas. The data regarding the digital economy and the coordinated development of urban and rural areas are derived from the "China Statistical Yearbook" published from 2012 to 2021. China's Digital Economy Innovation and Entrepreneurship Index, published by the Center for Enterprise Research at Peking University, provides access to variable data, including the number of new enterprises, foreign investment, venture capital, patents granted, trademark registrations, and software copyright registrations. The descriptive statistics of the variables are given in Table 2.

Table 2. Descriptive statistics of variables.

	Variables	Sample Size	Average Value	Standard Deviation	Minimum Value	Maximum Value
Explained variables	Urds	300	0.466	1.474	−0.257	8.723
Explanatory variables	Diec	300	0.797	0.758	0.067	4.078
Mediator variables	Urig	300	0.093	0.043	0.018	0.227
Control variables	Fepl Avsy Finance Scte	300 300 300 300	0.409 9.452 3.199 0.471	0.04 1.073 1.08 0.262	0.281 7.514 1.568 0.155	0.504 14.185 7.607 1.286

4.4. Model Setting

4.4.1. Panel Data Model

To test the validity of the research hypotheses, we first needed to develop the following fundamental model for the empirical examination of the direct impact mechanism that the digital economy has on coordinated urban–rural development:

$$\mathrm{Urds}_{i,t} = \alpha_0 + \alpha_1 \mathrm{Diec}_{i,t} + \alpha_c Z_{i,t} + \mu_i + \delta_t + \varepsilon_{i,t} \quad (3)$$

where i stands for the province code, t is time, Urds represents the level of the coordinated urban–rural development, Diec is the level of digital economy development, vector Z stands for a series of the control variables, μ represents the individual fixed effects of provinces that do not vary over time, δ represents the time fixed effects, and ε stands for the random disturbance term.

4.4.2. Spatial Econometric Model

Secondly, based on model (3), to discuss the spatial spillover effects of the digital economy on the coordinated development of urban and rural areas, we have used the SDM, the spatial autoregression model (SAR), and the spatial error model (SEM) for testing. The specific employed models are given below.

$$Urds_{i,t} = \alpha_0 + \rho WUrds_{i,t} + \phi_1 WDiec_{i,t} + \alpha_1 Diec_{i,t} + \phi_c WZ_{i,t} + \alpha_c Z_{i,t} + \mu_i + \delta_t + \varepsilon_{i,t} \quad (4)$$

$$Urds_{i,t} = \alpha_0 + \alpha_1 Diec_{i,t} + \alpha_c Z_{i,t} + \varepsilon_{i,t} \quad (5)$$

$$Urds_{i,t} = \alpha_0 + \alpha_1 Diec_{i,t} + \alpha_c Z_{i,t} + \rho WUrds_{i,t} + \mu_i + \delta_t + \varepsilon_{i,t} \quad (6)$$

where ρ represents the autoregressive regression coefficient, W is the spatial weight matrix, and ϕ_1 and ϕ_c stand for the spatial interaction terms of the core explanatory and the control variables, respectively. The connotations of Equations (5) and (6) are consistent with Equation (4).

4.4.3. Mediating Effect Model

The digital economy can impact coordinated urban–rural development by affecting the income gap between urban and rural residents. For the empirical analysis, a model of the mediating effect is developed, as shown in the Equations below:

$$M_{i,t} = \phi_0 + \phi_1 Diec_{i,t} + \phi_c Z_{i,t} + \mu_i + \delta_t + \varepsilon_{i,t} \quad (7)$$

$$Urds_{i,t} = \beta_0 + \beta_1 Diec_{i,t} + \beta_2 M_{i,t} + \beta_c Z_{i,t} + \mu_i + \delta_t + \varepsilon_{i,t} \quad (8)$$

where M represents the mediating variable, indicating the urban–rural income gap (Urig).

4.5. Setting of the Spatial Weighting Matrix

To determine the distance between the provinces, we have utilised the two spatial weight matrices given below. The Equation (9) is the adjacency matrix, which is relatively easy to construct. If there is a common boundary between two different provinces, then the final value is 1; otherwise, it is 0. The Equation (10) is the economic distance matrix, which represents the difference in the level of economic development between provinces, expressed as the absolute value of the subtraction of each province's GDP. These two weighting matrices are set as follows:

$$W_{ij} = \begin{cases} 1 & \text{Area i is adjacent to area j} \\ 0 & \text{Area i is not adjacent to area j} \end{cases} \quad (9)$$

$$W_{ij} = \begin{cases} 1/\left|\overline{G_{pi}} - \overline{G_{pj}}\right| & i \neq j \\ 0 & i = j \end{cases} \quad (10)$$

$\overline{Gpi}$ and $\overline{Gpj}$ represent the difference in economic income (GDP) between province i and province j, respectively, and the other symbols have the same connotation as in Equation (4).

5. Empirical Testing and Analysis

5.1. Baseline Regression Analysis

Table 3 displays the effects of the digital economy on the coordinated development of urban and rural areas in each region. The findings of the Hausman test suggest that a fixed-effects model is preferable to a random-effects model. In light of this, the fixed-effects model was also utilised to estimate the parameters of this study. Model (1) displays the baseline regression results without the inclusion of control variables, while models (2) through (5) display the baseline regression results with the increasing inclusion of control factors. The estimated coefficient of the digital economy development level (Diec) on the urban–rural coordination development level (Urds) was found to be significantly more positive with or without the inclusion of the control variables, while the size of the decidable coefficient R^2 remains largely consistent. This suggests that, as the digital economy develops, the level of urban–rural coordination development in each region increases too. These results support H1.

Table 3. Baseline regression results.

Variables	(1)	(2)	(3)	(4)	(5)
Diec	0.309 *** (0.085)	0.311 *** (0.086)	0.292 *** (0.082)	0.293 *** (0.082)	0.333 *** (0.089)
Fepl		−0.319 (1.337)	−0.404 (1.280)	−0.432 (1.286)	−0.903 (1.341)
Avsy			0.532 *** (0.108)	0.538 *** (0.109)	0.530 *** (0.110)
Finance				0.026 (0.088)	0.012 (0.089)
Scte					−0.323 (0.264)
Year	YES	YES	YES	YES	YES
Province	YES	YES	YES	YES	YES
N	300	300	300	300	300
R^2	0.933	0.933	0.934	0.939	0.939

Note: *** mean significant at the 1%, with standard errors in brackets.

5.2. Spatial Correlation Analysis

5.2.1. Global Spatial Correlation Analysis

The spatial analysis requires the existence of a spatial correlation between the research variables. Primarily, Moran's I index and the Geary index are utilised to determine whether a spatial correlation exists. In this study, Moran's I index was used to analyse the spatial association between the digital economy (Diec) level and the level of urban–rural development coordination (Urds). Table 4 displays the results of this experiment. Moran's I index of the digital economy and the coordinated urban–rural development were considered positive at the 1% level, and the two exhibit spatial clustering phenomena, satisfying the requirements for spatial model analysis.

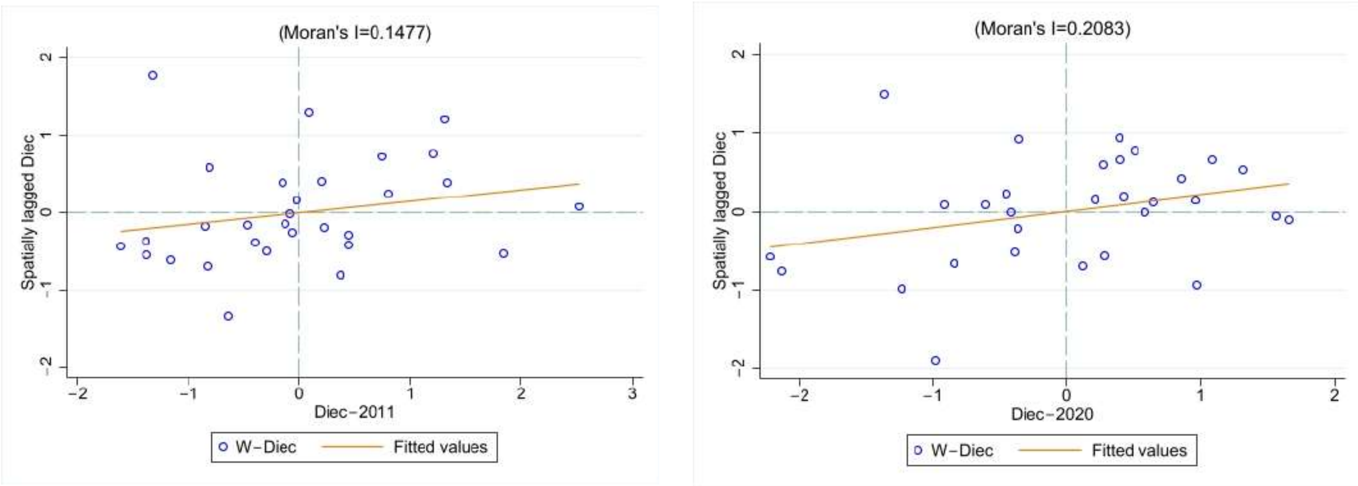
Table 4. Global Moran Index.

Year	Urds		Diec	
	Moran's I	Z Value	Moran's I	Z Value
2011	0.333 ***	5.029	0.226 ***	3.303
2012	0.331 ***	5.024	0.231 ***	3.212
2013	0.329 ***	5.012	0.165 *	2.363
2014	0.328 ***	4.999	0.161 *	2.361
2015	0.329 ***	4.986	0.168 *	2.458
2016	0.327 ***	4.961	0.192 **	2.716
2017	0.328 ***	4.939	0.218 **	2.985
2018	0.328 ***	4.909	0.151 *	2.046
2019	0.331 ***	4.829	0.061	1.045
2020	0.324 ***	4.564	0.013	0.517

Note: *, **, *** mean significant at the 5%, 1%, and 0.1% levels.

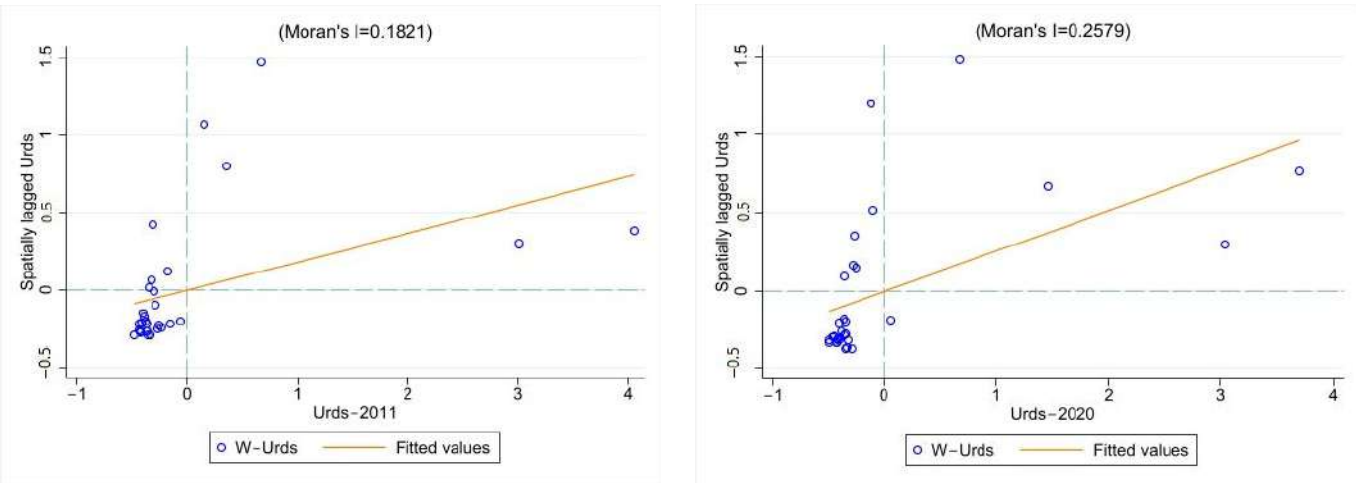
5.2.2. Local Spatial Correlation Analysis

The local Moran index offers a more precise depiction of the spatial correlation between the regions, and it also investigates the local spatial aggregation of the explored variables. In our study, the local Moran index was computed for each year from the sample, under the adjacency matrix. The results demonstrated a spatial association between the digital economy (Diec) and coordinated urban–rural growth (Urds). Figures 4 and 5 depict 2011 and 2020 Moran scatter plots for the digital economy and coordinated urban–rural development, respectively. As depicted in Figure 4, as the digital economy's level increased, many provinces fell into the first and third quadrants, with spatial spillover and diffusion effects gradually increasing. Most provinces were located in the first and third quadrants of Figure 5, demonstrating a spatial correlation between coordinated urban and rural development. We can conclude from the preceding study that there has been a significant spatial association between the digital economy and urban–rural development.



(**a**) (**b**)

Figure 4. Moran's scatter plot of the digital economy levels in 2011 and 2020. (**a**) is the Moran's scatter plot of the digital economy level in 2011, (**b**) is the Moran's scatter plot of the digital economy level in 2020.



(**a**) (**b**)

Figure 5. Moran's scatter plot of the level of urban–rural coordination in 2011 and 2020. (**a**) is the Moran's scatter plot of the level of urban–rural coordination in 2011, (**b**) is the Moran's scatter plot of the level of urban–rural coordination in 2020.

5.3. Analysis of the Spatial Econometric Estimation

In this study, we have employed the test idea of Elhorst (2015) [58], but we have also selected a suitable spatial econometric model using the four steps explained below.

Firstly, the Lagrange multiplier (LM) test results indicated that the choice of either the SEM or the SAR model was appropriate. Thus, the SDM that included both was to be selected. Secondly, the likelihood ratio (LR) test results indicated that the original hypothesis did not hold, indicating that the SDM model could not be degraded to the SAR model or the SEM model. Finally, the Hausman test results indicated that the choice of the fixed-effects model was more suitable for the estimation than the random-effects model.

The estimation results for the spatial Durbin, spatial lag, and the SEM, under the adjacency matrix, are given in Table 5.

Table 5. Regression results of the spatial model.

	(1) SDM	(2) SAR	(3) SEM
Diec_Main	0.299 *** (0.078)	0.260 *** (0.077)	0.174 ** (0.078)
Diec_Wx	0.627 *** (0.165)		
Diec_W-Direct	0.371 *** (0.087)	0.280 *** (0.084)	
Diec_W-Indirect	1.061 *** (0.241)	0.204 ** (0.080)	
Diec_W-Total	1.432 *** (0.298)	0.484 *** (0.152)	
rho	0.347 *** (0.078)	0.447 *** (0.070)	
lambda			0.441 *** (0.078)
sigma2_e	0.099 *** (0.008)	0.111 *** (0.009)	0.115 *** (0.010)
Year	YES	YES	YES
Province	YES	YES	YES
N	300	300	300
R^2	0.186	0.341	0.461

Note: **, *** mean significant at the 1%, and 0.1% levels.

To further examine the spatial spillover effects of the digital economy on coordinated urban–rural development, the effects of the explanatory variables of one region on the explained variables of the same and other regions were decomposed using the partial differential interpretation method into direct effects, indirect effects, and total effects [59]. The results are shown in Table 5.

Table 5 displays the results of the fixed-effects spatial models, which reveal that the signs and the numerical magnitudes of the regression coefficients of the variables in the SDM, SAR, and SEM models were found to be essentially consistent, with the results being highly credible. This empirical analysis reveals that all factors of the digital economy have had a considerable positive impact on the level of coordinated urban–rural development in neighbouring regions. This then indicates that the digital economy did in fact promote the level of coordinated urban–rural development in the neighbouring regions through the spatial spillover effects, thus providing support for H2. As far as explanatory variables go, the direct effect, indirect effect (the spatial spillover effect), and the total effect of the digital economy (Diec) all had a significant positive effect, indicating that the digital economy not only increased the coordinated urban–rural development level in the region but also improved it in neighbouring regions. This finding offers additional support for H2.

5.4. Further Research: Regional Heterogeneity

5.4.1. Regional Heterogeneity of the Direct Effects

Using the statistical system as well as the classification standards most recently released by the National Bureau of Statistics as a guide, in this study, we have also divided China’s 30 provinces (the municipalities directly under the Central Government and the autonomous regions) into three major regions: the eastern, the central, and the western regions. The regressions for each region used the SDM to experimentally examine regional variability in the digital economy's direct effects and spatial spillover effects on China's coordinated growth of urban and rural areas. Table 6 shows that the digital economy positively influences eastern and central regions' coordinated urban–rural growth. On the other hand, a negative coefficient was found for its effect in the western region. In other words, these results indicate that the digital economy did significantly promote coordinated urban–rural development in the eastern region and that it did significantly inhibit this development in the western region, thereby providing support for H3. Due to the small sample size, the promotion effect in the central region was not found to be statistically significant.

Table 6. Regional heterogeneity test.

	(1) East	(2) Middle	(3) West
Direct-Diec	0.364 **	0.046	−0.015 *
	(0.150)	(0.042)	(0.009)
Spatial-Diec	1.462 ***	0.019	0.062 *
	(0.306)	(0.087)	(0.035)
rho	0.438 ***	0.409 ***	−0.817 ***
	(0.114)	(0.137)	(0.173)
sigma2_e	0.107 ***	0.002 ***	0.000 ***
	(0.015)	(0.000)	(0.000)
Year	YES	YES	YES
Province	YES	YES	YES
N	110	80	110
R^2	0.381	0.354	0.384

Note: *, **, *** mean significant at the 5%, 1%, and 0.1% levels, with standard errors in brackets.

5.4.2. Regional Heterogeneity of the Spatial Effects

If the regional heterogeneity is analysed from the perspective of the direct effects alone, the results could be biased. Table 6 additionally includes the spatial spillover effects, which should be employed in further investigations of the regional heterogeneity and the digital economy, in relation to the coordinated urban–rural development from a spatial spillover perspective. Table 6 further demonstrates that, at the national level, the spatial spillover effects of the digital economy on coordinated urban–rural development were significantly positive.

From the combined effects, the digital economy was shown to have a significant positive effect on the coordinated urban–rural development level in the eastern region through the direct and spatial spillover effects. In the central region, however, the direct and the spatial spillover effects of the digital economy on the coordinated urban–rural development level were not found to be significant, but they rather exhibited a general tendency to improve it. Finally, in the western region, the direct and the spatial promotion effects were also not significant, even though they showed a general tendency to increase. Additionally, in the western region, the direct and the spatial promotion effects of the digital economy were more pronounced, but they did exhibit two opposing effects: direct inhibition and spatial promotion.

Possible explanations for the differences between the direct and the spatial spillover effects include the fact that the spatial spillover effects exhibit different regional trends due to the factors such as network structure, knowledge gaps and absorptive capacity, and economic and policy environments. The central and the western regions are still in the initial development stage of the digital economy due to a lack of digital infrastructure, digital talents, and other resources, while the eastern regions have developed the digital economy earlier and consequently have a higher degree of development and aggregation level than the central regions, allowing them to reap the benefits of the digital economy.

5.5. The Mediating Effect of the Digital Economy on Coordinated Urban–Rural Development

Following Fritz and MacKinnon's research (2007) [60], a mechanism analysis was first conducted to investigate the impact of the digital economy on the income gap between urban and rural residents. Model (1) in Table 7 shows regression estimates from a fixed-effects model. The results demonstrate that the digital economy reduces the urban–rural income gap. This study also employed the mediating effects model to empirically investigate the relationship between digital economy development and coordinated urban–rural development by examining the income gap between urban and rural areas. Combining the regression results of models (2) and (3) in Table 7, the estimated coefficients of the core explanatory variables and the mediating variables were found to be significant at the 1% level, and therefore, no further Sobel test was required. Furthermore, there is a mediating effect with the income gap between urban and rural residents (Urig) as the mediating variable, which supports H4.

Table 7. Mediation effect test.

Variables	(1) Urig	(2) Urds	(3) Urds
Diec	−0.23 *** (0.003)	0.371 *** (0.079)	0.214 *** (2.16)
Urig			−6.868 *** (1.421)
Control variables	YES	YES	YES
Year	YES	YES	YES
Province	YES	YES	YES
R^2	0.383	0.664	0.688
N	300	300	300

Note: *** mean significant at the 1%, with standard errors in brackets.

5.6. Robustness Tests

To ensure the consistency and stability of the empirical results, we used the following three methods. The results of applying a 1% tail-shrinking to the core explanatory variables are shown in model (1) of Table 8, and they represent the firstly employed method. The second method that was employed included replacing the core explanatory variables and reconstructing the digital economy level for the regression, in accordance with Tao's (2020) research [61], and the regression results of model (2) are presented in Table 8. The third employed method included the replacement of the spatial weight matrix with an economic distance matrix. The resulting model (3) is presented in Table 8 as well. The estimation results of all the models shown in Table 8 indicate that the core explanatory variables are significantly positive, though with different levels of confidence, except for variations in the estimated values. This shows that the found empirical results are more robust.

Table 8. Robustness tests.

	(1)	(2)	(3)
Diec	0.352 *** (0.078)	7.329 *** (1.223)	0.138 ** (0.068)
W-Diec	0.544 ***	13.013 ***	0.495 ***
rho	0.330 *** (0.079)	0.179 ** (0.083)	0.382 *** (0.102)
sigma2_e	0.102 *** (0.008)	0.410 *** (0.034)	0.258 *** (0.022)
N	300	300	300
R^2	0.188	0.179	0.285

Note: **, *** mean significant at the 1%, and 0.1% levels, with standard errors in brackets.

6. Conclusions

The rapid growth of the digital economy has made it a key factor in the high-quality development of China's economy, with digitalisation and artificial intelligence seen as the future economic development trend. Based on the balanced panel data obtained for 30 provinces (the municipalities directly under the Central Government and the autonomous regions) in China for the period from 2011 to 2020, this paper deals with the effect of the digital economy on coordinated urban–rural development, using a combination of panel fixed-effects models, the mediating effects model, and the SDM. Conclusions that can be drawn based on the results of our analysis are discussed below.

First, the results of the benchmark regression indicate that the development of the digital economy has significantly reduced the dual economic structure of urban and rural areas and that it has fostered the growth of coordinated urban–rural development. Second, the results of the SDM stipulate that the existence of a significant positive spatial spillover effect of the digital economy on coordinated urban–rural development is present and that the found results were still significant under the transformation of the economic distance matrix. These results are found to be highly robust. Third, the digital economy affects urban–rural coordinated development by reducing the urban–rural income gap. Fourth, the results of the heterogeneity test point out that the positive impact of the digital economy on coordinated urban–rural development is robust as well. Finally, the results of the heterogeneity test show that the impact of the digital economy on coordinated regional development is regionally heterogeneous, where the digital economy has a significant positive effect on urban and rural development in the eastern region, a non-significant positive effect in the central region, and a significant inhibiting effect in the western region. In terms of the spatial spillover effects, the digital economy has exhibited a positive spillover effect on the coordinated development of the urban and rural areas in the eastern region, whereas it has no promotion effect on the central and western regions. In summary, the digital economy innovation dividend was found to be significantly higher in the eastern region than in the central and western regions.

7. Policy Recommendations and Limitations

7.1. Policy Recommendations

In response to these findings, the following policy recommendations are presented:

(1) The simultaneous development of the digital economy in urban and rural areas should be promoted. Moreover, what should also be promoted is the integrated development of the urban and rural areas, and the digital economy's dividends should be fully released. To further explain, firstly, the application of digital technology in rural areas needs to be strengthened, parallel with the act of active promotion of the application of new agricultural development models based on artificial intelligence, the Internet of Things, big data, and 5G technology. What also needs to be empowered is the development of digital villages with digital technology. Secondly, what should be accelerated is the construction of an intelligent agricultural production system, the integration of agricultural and rural data, the development of the existing agricultural information service platforms, and the enhancement of agricultural information service capabilities. This should serve to establish modern agriculture in the countryside with strong and enduring competitive advantages. Thirdly, investment in the education and training of farmers should be increased, with an emphasis put on the development of their digital literacy and vocational skills, as farmers' wages and incomes are significantly influenced by their level of knowledge and proficiency. Finally, increasing the knowledge and skills of farmers will narrow the gap between the labour skill endowment of the urban and the rural workforce. This will thereby enhance their employment competitiveness and ensure the stability and sustainability of employment for rural residents.

(2) Given the heterogeneity of the digital economy development between regions, cities, and rural areas and groups, it seems to be necessary to formulate differentiated and

hierarchical digital economy development strategies. Firstly, we should promote the construction of the "East is Digital, West is Digital" and "Broadband China" projects, as well as the construction and the layout of rural digital infrastructure. Secondly, the "New Infrastructure" program should increase their investment in rural areas and thusly gradually improve the digital infrastructure environment and the digital technology penetration rate in rural areas. In turn, this will reduce the cost of searching and absorbing information in rural areas and additionally narrow the "digital divide". What needs to be carried out, thirdly, for the coordinated development of the urban and rural areas is to stimulate the endogenous momentum of the digital transformation of the traditional industries, promote the stable development of the digital economy, and further consolidate the dividend effect caused by digital technology. To do this, the government should play its role of guidance and support, thus leading the digital transformation of the traditional industries in a reasonable manner, as well as providing certain financial and tax policy support.

(3) Another segment that asks for action is the full utilisation of the digital economy's spatial spillover effect on the coordinated development of the urban and rural areas, as well as the information radiation-driven effect of the relatively developed digital economies in the surrounding areas. We should promote the rationalisation of the layout of the digital economy industry together with the even distribution of digital resources. We should also direct the spatial concentration of the digital economy to rural areas, alleviate the contradiction between the resources, the environment, and the development in rural areas, and, at last, narrow the "digital divide" between the urban and rural areas. In that way, we would be promoting the coordinated development of the urban and rural areas. What should be strengthened is cross-regional exchanges and cooperation, where governments should actively build cross-regional cooperation platforms, promote and support inter-regional cooperation and exchanges, and, in that way, create a good environment for cooperation and innovation. Governments should finally promote the reasonable flow of talents, capital, and other elements across the regions to be able to build sharing practices of urban and rural resources and propel the development of underdeveloped rural areas more effectively.

7.2. Limitations and Prospects

This study includes the digital economy and coordinated urban–rural development in the research framework, examines the impact and mechanism of the digital economy on coordinated urban–rural development from the perspective of urban–rural income disparity as well as spatial spillover, and puts forward policy recommendations for promoting coordinated urban–rural development. However, there are several limitations of the study.

First, there are many factors affecting coordinated urban–rural development, and this study measures the level of coordinated urban–rural development from the perspective of economic structure, which may not provide a comprehensive measure of urban and rural development.

Second, the sample used in this study is based on provincial-level data, which may bias the results to a certain extent due to the small sample size; using prefectural or county-level data would be more detailed and accurate.

Finally, this study has only looked at the current coordinated urban–rural development of the digital economy, and the driving effect of the digital economy on urban–rural development in the long term may yield different results.

Future studies could use more refined measures of coordinated urban–rural development and more detailed data and methods such as dynamic modelling to explore the long-term effects of the digital economy on urban–rural development.

Author Contributions: Conceptualisation and writing, Y.C. and D.Z.; review and editing, Y.C.; methodology, D.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant no. 72164002, 71762003).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors are grateful to the editors and the anonymous referees for their constructive and thorough comments, which contributed to the improvement of this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, C.; LeGates, R.; Zhao, M.; Fang, C. The changing rural-urban divide in China's megacities. *Cities* **2018**, *81*, 81–90. [CrossRef]
2. Fan, J.; Zhao, H.; Guo, R. The New Trend and Coping Strategies of Regional Development Gap in China. *Econ. Geogr.* **2022**, *42*, 1–11.
3. Liu, Y.; Zang, Y.; Yang, Y. China's rural revitalization and development: Theory, technology and management. *J. Geogr. Sci.* **2020**, *30*, 1923–1942. [CrossRef]
4. Wang, H.; Chen, L. Review on the development theories of rural and urban areas in West. *Econ. Geogr.* **2006**, *26*, 463–468.
5. Lysgard, H.K. The assemblage of culture-led policies in small towns and rural communities. *Geoforum* **2019**, *101*, 10–17. [CrossRef]
6. Ma, L.; Liu, S.; Fang, F.; Che, X.; Chen, M. Evaluation of urban-rural difference and integration based on quality of life. *Sustain. Cities Soc.* **2020**, *54*, 101877. [CrossRef]
7. Yang, Y.; Bao, W.; Wang, Y.; Liu, Y. Measurement of urban-rural integration level and its spatial differentiation in China in the new century. *Habitat Int.* **2021**, *117*, 102420. [CrossRef]
8. Caffyn, A.; Dahlstrom, M. Urban-rural interdependencies: Joining up policy in practice. *Reg. Stud.* **2005**, *39*, 283–296. [CrossRef]
9. Shunan, W.; Jiansheng, C. The Techno-Economic Paradigm of the Digital Economy. *Shanghai J. Econ.* **2019**, *38*, 80–94.
10. Leng, X. Digital revolution and rural family income: Evidence from China. *J. Rural Stud.* **2022**, *94*, 336–343. [CrossRef]
11. Marino, A.; Pariso, P. Digital government platforms: Issues and actions in Europe during pandemic time. *Entrep. Sustain. Issues* **2021**, *9*, 462. [CrossRef]
12. Marino, A.; Pariso, P.; Picariello, M. Transition towards the artificial intelligence via re-engineering of digital platforms: Comparing European Member States. *Entrep. Sustain. Issues* **2022**, *9*, 350. [CrossRef] [PubMed]
13. Bauer, J.M. The Internet and income inequality: Socio-economic challenges in a hyperconnected society. *Telecommun. Policy* **2018**, *42*, 333–343. [CrossRef]
14. Ding, C.; Liu, C.; Zheng, C.; Li, F. Digital Economy, Technological Innovation and High-Quality Economic Development: Based on Spatial Effect and Mediation Effect. *Sustainability* **2022**, *14*, 216. [CrossRef]
15. Ma, D.; Zhu, Q. Innovation in emerging economies: Research on the digital economy driving high-quality green development. *J. Bus. Res.* **2022**, *145*, 801–813. [CrossRef]
16. Kristoffersen, E.; Blomsma, F.; Mikalef, P.; Li, J. The smart circular economy: A digital-enabled circular strategies framework for manufacturing companies. *J. Bus. Res.* **2020**, *120*, 241–261. [CrossRef]
17. Liu, Q.; Trevisan, A.H.; Yang, M.; Mascarenhas, J. A framework of digital technologies for the circular economy: Digital functions and mechanisms. *Bus. Strategy Environ.* **2022**, *31*, 2171–2192. [CrossRef]
18. Wang, X.; Sun, X.; Zhang, H.; Xue, C. Digital Economy Development and Urban Green Innovation CA-Pability: Based on Panel Data of 274 Prefecture-Level Cities in China. *Sustainability* **2022**, *14*, 2921. [CrossRef]
19. Zhang, J.; Lyu, Y.; Li, Y.; Geng, Y. Digital economy: An innovation driving factor for low-carbon development. *Environ. Impact Assess. Rev.* **2022**, *96*, 106822. [CrossRef]
20. Guan, H.; Guo, B.; Zhang, J. Study on the Impact of the Digital Economy on the Upgrading of Industrial Structures-Empirical Analysis Based on Cities in China. *Sustainability* **2022**, *14*, 11378. [CrossRef]
21. Liu, Y.; Yang, Y.; Li, H.; Zhong, K. Digital Economy Development, Industrial Structure Upgrading and Green Total Factor Productivity: Empirical Evidence from China's Cities. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2414. [CrossRef] [PubMed]
22. Zhavoronok, A.; Popelo, O.; Shchur, R.; Ostrovska, N.; Kordzaia, N. The role of digital technologies in the transformation of regional models of households' financial behavior in the conditions of the national innovative economy development. *Ing. Syst. D'inf.* **2022**, *27*, 613–620. [CrossRef]
23. Popelo, O.; Tulchynska, S.; Krasovska, G.; Kostiunik, O.; Raicheva, L.; Mykhalchenko, O. The Impact of the National Economy Digitalization on the Efficiency of the Logistics Activities Management of the Enterprise in the Conditions of Intensifying International Competition. *J. Theor. Appl. Inf. Technol.* **2023**, *101*, 123–134.
24. Myovella, G.; Karacuka, M.; Haucap, J. Digitalization and economic growth: A comparative analysis of Sub-Saharan Africa and OECD economies. *Telecommun. Policy* **2020**, *44*, 101856. [CrossRef]
25. de Silva, I. Tackling the Challenges Raised by the Digitalization of the Economy: Recent Experiences of the French Competition Authority. *Antitrust Bull.* **2019**, *64*, 3–10. [CrossRef]

26. Pan, W.; Xie, T.; Wang, Z.; Ma, L. Digital economy: An innovation driver for total factor productivity. *J. Bus. Res.* **2022**, *139*, 303–311. [CrossRef]
27. Zhao, H.; Zheng, X.; Yang, L. Does digital inclusive finance narrow the urban-rural income gap through primary distribution and redistribution? *Sustainability* **2022**, *14*, 2120. [CrossRef]
28. Athique, A. Integrated commodities in the digital economy. *Media Cult. Soc.* **2020**, *42*, 554–570. [CrossRef]
29. He, J.; Peng, J.; Zeng, G. The Spatiality of the Creative Digital Economy: Local Amenities to the Spatial Agglomeration of Creative E-Freelancers in China. *J. Knowl. Econ.* **2022**, *13*, 1–22. [CrossRef]
30. Rani, U.; Furrer, M. Digital labour platforms and new forms of flexible work in developing countries: Algorithmic management of work and workers. *Compet. Chang.* **2021**, *25*, 212–236. [CrossRef]
31. Zhou, C.; Zhang, D.; Chen, Y. Theoretical framework and research prospect of the impact of China's digital economic development on population. *Front. Earth Sci.* **2022**, *10*, 988608. [CrossRef]
32. Karar, H. Algorithmic Capitalism and the Digital Divide in Sub-Saharan Africa. *J. Dev. Soc.* **2019**, *35*, 514–537. [CrossRef]
33. Sujarwoto, S.; Tampubolon, G. Spatial inequality and the Internet divide in Indonesia 2010–2012. *Telecommun. Policy* **2016**, *40*, 602–616. [CrossRef]
34. Jun, L. An empirical study on the urban-rural digital divide that continues to widen the urban-rural income gap. *Stat. Decis.* **2017**, *33*, 119–121.
35. Yaping, H.; Canning, X. The impact of the Internet on the urban-rural income gap:A test based on Chinese facts. *Econ. Surv.* **2019**, *36*, 25–32.
36. Fan, Y.; Xu, H.; Ma, L. Characteristics and Mechanism Analysis of the Influence of Digital Economy on the Income Gap between Urban and Rural Residents. *China Soft Sci.* **2022**, *37*, 181–192.
37. Jiang, Q.; Li, Y.; Si, H. Digital Economy Development and the Urban-Rural Income Gap: Intensifying or Reducing. *Land* **2022**, *11*, 1980. [CrossRef]
38. Wolcott, E.L. Employment inequality: Why do the low-skilled work less now? *J. Monet. Econ.* **2021**, *118*, 161–177. [CrossRef]
39. Deichmann, U.; Goyal, A.; Mishra, D. Will digital technologies transform agriculture in developing countries? *Agric. Econ.* **2016**, *47*, 21–33. [CrossRef]
40. Islamutdinov, V.F. Institutional change within the context of digital economy. *J. Inst. Stud.* **2020**, *12*, 142–156. [CrossRef]
41. Chen, W.; Wang, Q.; Zhou, H. Digital Rural Construction and Farmers' Income Growth: Theoretical Mechanism and Micro Experience Based on Data from China. *Sustainability* **2022**, *14*, 11679. [CrossRef]
42. Pfeiffer, J.; Gabriel, A.; Gandorfer, M. Understanding the public attitudinal acceptance of digital farming technologies: A nation-wide survey in Germany. *Agric. Hum. Values* **2021**, *38*, 107–128. [CrossRef]
43. Tobler, W.R. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **1970**, *46* (Suppl. S1), 234–240. [CrossRef]
44. Cong, X.; Wang, S.; Wang, L.; Saparauskas, J.; Gorecki, J.; Skibniewski, M.J. Allocation Efficiency Measurement and Spatio-Temporal Differences Analysis of Digital Infrastructure: The Case of China's Shandong Province. *Systems* **2022**, *10*, 205. [CrossRef]
45. Zhou, A. Digital infrastructure and economic growth—Evidence for China. *J. Infrastruct. Policy Dev.* **2022**, *6*, 1397. [CrossRef]
46. Li, Z.; Liu, Y. Research on the Spatial Distribution Pattern and Influencing Factors of Digital Economy Development in China. *IEEE Access* **2021**, *9*, 63094–63106. [CrossRef]
47. Du, M.; Huang, Y.; Dong, H.; Zhou, X.; Wang, Y. The measurement, sources of variation, and factors influencing the coupled and coordinated development of rural revitalization and digital economy in China. *PLoS ONE* **2022**, *17*, e0277910. [CrossRef] [PubMed]
48. Li, Z.; Liu, C.; Chen, X. Power of Digital Economy to Drive Urban-Rural Integration: Intrinsic Mechanism and Spatial Effect, from Perspective of Multidimensional Integration. *Int. J. Environ. Res. Public Health* **2022**, *19*, 15459. [CrossRef]
49. Li, Y.X.; Li, Z.X. Digital economy development and urban-rural income gap in the context of rural revitalization. *Financ. Econ.* **2022**, *43*, 60–67+96.
50. Yin, Z.H.; Choi, C.H. Does e-commerce narrow the urban-rural income gap? Evidence from Chinese provinces. *Internet Res.* **2022**, *32*, 1427–1452. [CrossRef]
51. Nanna, L. The digital economy, the urban-rural income gap and economic development resilience. *J. Tech. Econ. Manag.* **2022**, *43*, 10–14.
52. Jun, W.; Huatang, X. Does the growth of the digital economy narrow the income gap between urban and rural residents? *Reform Econ. Syst.* **2021**, *39*, 56–61.
53. Ma, Q.; Tariq, M.; Mahmood, H.; Khan, Z. The nexus between digital economy and carbon dioxide emissions in China: The moderating role of investments in research and development. *Technol. Soc.* **2022**, *68*, 101910. [CrossRef]
54. He, L.; Zhang, X.L. The distribution effect of urbanization: Theoretical deduction and evidence from China. *Habitat Int.* **2022**, *123*, 102544. [CrossRef]
55. Guo, S.L.; Wang, B.B.; Zhou, K.; Wang, H.; Zeng, Q.P.; Xu, D.D. Impact of Fiscal Expenditure on Farmers' Livelihood Capital in the Ethnic Minority Mountainous Region of Sichuan, China. *Agriculture* **2022**, *12*, 881. [CrossRef]
56. Le, H.T.; Booth, A.L. Inequality in V ietnamese Urban–Rural Living Standards, 1993–2006. *Rev. Income Wealth* **2014**, *60*, 862–886.
57. Su, C.W.; Song, Y.; Ma, Y.T.; Tao, R. Is financial development narrowing the urban-rural income gap? A cross-regional study of China. *Pap. Reg. Sci.* **2019**, *98*, 1779–1800. [CrossRef]

58. Halleck Vega, S.; Elhorst, J.P. The SLX model. *J. Reg. Sci.* **2015**, *55*, 339–363. [CrossRef]
59. LeSage, J.; Pace, R.K. *Introduction to Spatial Econometrics*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2009.
60. Fritz, M.S.; MacKinnon, D.P. Required sample size to detect the mediated effect. *Psychol. Sci.* **2007**, *18*, 233–239. [CrossRef]
61. Tao, Z.; Zhi, Z.; Shangkun, L. Digital Economy, Entrepreneurial Activity and High Quality Development-Empirical Evidence from Chinese Cities. *J. Manag. World* **2020**, *36*, 65–76.

sustainability

Article

Smarter Sustainable Tourism: Data-Driven Multi-Perspective Parameter Discovery for Autonomous Design and Operations

Raniah Alsahafi [1], Ahmed Alzahrani [1] and Rashid Mehmood [2,*]

[1] Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[2] High-Performance Computing Center, King Abdulaziz University, Jeddah 21589, Saudi Arabia
* Correspondence: rmehmood@kau.edu.sa

Abstract: Global natural and manmade events are exposing the fragility of the tourism industry and its impact on the global economy. Prior to the COVID-19 pandemic, tourism contributed 10.3% to the global GDP and employed 333 million people but saw a significant decline due to the pandemic. Sustainable and smart tourism requires collaboration from all stakeholders and a comprehensive understanding of global and local issues to drive responsible and innovative growth in the sector. This paper presents an approach for leveraging big data and deep learning to discover holistic, multi-perspective (e.g., local, cultural, national, and international), and objective information on a subject. Specifically, we develop a machine learning pipeline to extract parameters from the academic literature and public opinions on Twitter, providing a unique and comprehensive view of the industry from both academic and public perspectives. The academic-view dataset was created from the Scopus database and contains 156,759 research articles from 2000 to 2022, which were modelled to identify 33 distinct parameters in 4 categories: Tourism Types, Planning, Challenges, and Media and Technologies. A Twitter dataset of 485,813 tweets was collected over 18 months from March 2021 to August 2022 to showcase the public perception of tourism in Saudi Arabia, which was modelled to reveal 13 parameters categorized into two broader sets: Tourist Attractions and Tourism Services. The paper also presents a comprehensive knowledge structure and literature review of the tourism sector based on over 250 research articles. Discovering system parameters are required to embed autonomous capabilities in systems and for decision-making and problem-solving during system design and operations. The work presented in this paper has significant theoretical and practical implications in that it improves AI-based information discovery by extending the use of scientific literature, Twitter, and other sources for autonomous, holistic, dynamic optimizations of systems, promoting novel research in the tourism sector and contributing to the development of smart and sustainable societies.

Keywords: smart tourism; sustainable tourism; natural language processing (NLP); big data analytics; deep learning; machine learning; unsupervised learning; Bidirectional Encoder Representations from Transformers (BERT); literature review; smart societies

Citation: Alsahafi, R.; Alzahrani, A.; Mehmood, R. Smarter Sustainable Tourism: Data-Driven Multi-Perspective Parameter Discovery for Autonomous Design and Operations. *Sustainability* **2023**, *15*, 4166. https://doi.org/10.3390/su15054166

Academic Editor: Jun (Justin) Li

Received: 1 February 2023
Revised: 19 February 2023
Accepted: 22 February 2023
Published: 25 February 2023

1. Introduction

Smart tourism [1,2] has developed as a sector of the smart societies concept [3,4], which integrates traditional tourism practices with the use of smart technologies to offer transformative and tailored solutions for the specific needs and requirements of travelers. However, the exploitative practices of capitalist approaches have had a detrimental effect on social, environmental, and economic sustainability. In response, sustainable tourism has emerged as an alternative approach that seeks to enhance both the tourism experience and promote longer-term sustainability [5–10].

1.1. Tourism in the Global Economy

Prior to COVID-19 (2019), the tourism sector, which was nearly three times larger than agriculture [11], amounted to 10.3% of the world's GDP, approximately USD 9.6 trillion [12]. In 2019, international visitors spent approximately USD 1.8 trillion, which represented 6.8% of total exports. In the same year, the sector accounted for one-quarter of all new jobs created globally and made up 10.3% of all jobs totaling 333 million [12].

COVID-19 severely impacted the tourism sector, reducing its contributions to the global GDP by nearly half in 2020, and subsequently recovering slowly in 2021 to a total global GDP contribution of USD 5.8 trillion [13]. In 2020, 62 million jobs were lost in the tourism sector, and the spending by domestic and international visitors declined by nearly 50% and 70%, respectively [12].

1.2. A Case Study on Tourism in Saudi Arabia: Diversification of Oil-Based Economy

For Saudi Arabia, in 2019, prior to COVID-19, travel and tourism amounted to 9.7% of the total national GDP, equaling 77.8 billion USD (see "Saudi Arabia 2022 Annual Research: Key highlights" at [12]). The sector contributed 1.58 million jobs, making up a 12.2% share of national jobs. The spending by domestic and international visitors in Saudi Arabia was 16.6 billion USD and 29.9 billion USD (10.4 percent of total exports), respectively.

There was a significant drop in all these economic figures due to the pandemic in Saudi Arabia. The 2021 economic figures for travel and tourism are as follows. The tourism sector was 6.5% (down from 9.7%) of the total national GDP, equaling 51.5 billion USD, and contributed 1.3 million jobs, making up a 10% share of national jobs. The spending by domestic and international visitors in Saudi Arabia was 16.2 billion USD and 6.1 billion USD (2.3% of total exports), respectively. The relatively low decline in domestic spending across the pandemic time is because a large part of local tourism is based on rural, natural, or religious sites and activities. Moreover, the restrictions on international travel and tourism were compensated by domestic travel. These economic figures for Saudi Arabia and other countries can be found in [12].

These numbers clearly show the vulnerability of the tourism sector and the national and global economies.

It is high time for new investors and other stakeholders in the tourism sector in Saudi Arabia, explained as follows. The Saudi economy has been heavily dependent on the oil industry. Saudi Arabia's vast oil reserves and position as the world's largest exporter of oil give it significant economic leverage in the global market. Its role in OPEC and close economic relationships with major oil-consuming countries allow it to influence the price and stability of the oil, making it a key player in the global geoeconomics of oil.

Saudi Arabia is diversifying its economy to reduce its reliance on oil, create new job opportunities, attract foreign investment, and prepare for the future in order to provide more stability, be more attractive to investors, and ensure its long-term prosperity. For instance, Saudi Arabia is investing in renewable energy as part of its efforts to diversify its energy mix and reduce reliance on oil. The goal is to produce 50% of the nation's electricity from clean energy sources by 2030, and the country has made significant progress in the development of solar and wind energy. It is also exploring other forms of renewable energy, such as geothermal and tidal energy, and has established initiatives and programs to support the development of renewables.

Tourism is an important target sector in Saudi Arabia's efforts to diversify its economy [14]. It is expected that the tourism sector will become the white oil, i.e., a major source of revenue for Saudi Arabia, similar to how (black) oil has traditionally been a major contributor to the country's economy [15]. Saudi Arabia, which is located in the Middle East, is a country known for its rich cultural history and significant religious sites such as Mecca and Medina. The government has been working to develop the country's tourism industry in recent years, with a focus on both domestic and international tourism. To support this growth, the government has implemented initiatives, for instance, the development of new infrastructure, such as hotels and airports, and the promotion of the country as a destination for events, conferences, and other business activities. There are opportunities for growth in various areas of tourism, including cultural and religious tourism, adventure and nature-based tourism, business and conference tourism, and medical tourism. In addition, Saudi Arabia is home to a number of cultural and natural attractions, including museums, galleries, the Asir National Park, and the Al-Hasa Oasis. Its cities, including Riyadh and Jeddah, are also popular tourist destinations. The planned city of NEOM is being developed as a hub for advanced technology and innovation, and there are several other planned tourism developments in Saudi Arabia, including Qiddiyah, Amaala, Trojena, Oxagon, The Line, Sindalah, and the Red Sea Project. Some of these are connected to NEOM city [16].

As Saudi Arabia and other Gulf countries compete in the tourism industry, they each offer unique cultural and natural attractions, modern infrastructure and facilities, and a convenient location for tourists from around the world. All of these recent developments and factors make researching the tourism sector in Saudi Arabia exciting and lucrative.

1.3. Technologies and Their Impact on Tourism

The tourism industry is being transformed by a number of key technologies, including virtual and augmented reality, artificial intelligence, machine and deep learning, the Internet of Things (IoT), mobile technologies, social media, big data and analytics, and blockchain. The IoT is expected to positively impact the tourism industry by providing timely or real-time interactions with tourists in areas such as transportation, attractions, tours, shopping, and hotels [17]. Artificial intelligence and machine learning are being employed to improve the customer experience through personalized recommendations and helping with trip planning, finding the nearest restaurants, suggesting road conditions, and many other applications for reservations, hotels, transport, and restaurants [18]. Virtual and augmented reality is being used to create immersive experiences for travelers [19]. Mobile technologies and social media have made it easier for travelers to book trips, find things to do, and connect with other travelers [20]. Big data and analytics are helping the industry to understand traveler behavior and preferences [21]. Blockchain has the potential to revolutionize the way the industry operates by enabling secure, transparent, and decentralized transactions [22]. These technologies are coming together to provide innovative capabilities and services including forecasting tourism demand [23], mining tourist locations, pathways, and travel itineraries [24], the visualization of tourist mobility activity patterns, identifying tourist congestion zones and tourist routes in specific areas [25], measuring tourist satisfaction [26], support for decision-making [27], identifying factors related to tourist satisfaction [28], integration of the Android platform and GPS services for guiding tourists to their preferred sites [29], augmented reality for historical tourism using mobile and smart devices [30,31], and more.

These technologies are all playing a significant role in transforming the tourism industry into smart tourism and shaping the way we travel in the future. Smart tourism [1,2] has developed as a sector of the smart societies concept [3,4], which integrates traditional tourism practices with the use of smart technology to offer tailored solutions for the specific needs and requirements of travelers.

1.4. The Need for Smart, Responsible, and Sustainable Tourism

Tourism is a multi-faceted industry that can bring economic and cultural benefits, but it can also pose challenges [32–34]. Environmental sustainability is a major concern, as the growth of tourism can put a strain on natural resources and ecosystems and contribute to climate change [35]. Overcrowding and over-tourism can lead to social and environmental impacts and overburden local infrastructure [36–40]. The economic impact can be positive, but can also lead to disruption and inequality [41]. Cultural sensitivity is also important as the influx of tourists can lead to cultural tension [37]. Health and safety risks, such as the spread of disease, crimes, and natural disasters, are also a concern [38,42]. Geopolitical dynamics and wars are also worsening the situation [43,44]. Many innovative and radical technologies are emerging frequently [45,46] while the existing technologies are evolving at a fast pace, and these advancements and dynamics are making unforeseeable impacts on societies and hence on tourism, its nature, and requirements [30,31,47]. It is important for all stakeholders, such as governments, businesses, and local communities, to work together to develop sustainable and responsible tourism practices, a concept being pursued internationally under the sustainable tourism umbrella [5–10].

A holistic and dynamic (interactive, timely, or real-time) understanding of the sector and related global, regional, national, and cultural issues is needed to drive innovation and improvements toward smart and sustainable tourism [42]. Further details about the research gap can be found in Section 2, and a discussion on the novelty and utilization of this work can be found in Section 6.

1.5. This Work

This study uses our data-driven approach to model the tourism industry through the lens of both academics and the general public, using 156,000 research papers and 485,000 tweets. By combining advanced technologies such as deep learning and big data, we have developed a machine learning pipeline to extract parameters from both the academic literature and public opinions on Twitter. This approach gives a unique and comprehensive view of the tourism industry from two differing perspectives. These perspectives are not isolated and have some impact on each other, but they still have distinct and significant variations.

The aim of this study is to gain a comprehensive understanding of tourism, drive future research through cutting-edge technologies, and ultimately develop a theory and approach for smarter tourism that supports sustainable future societies. The paper presents a comprehensive knowledge structure and literature review of the tourism sector, drawing on more than 250 research articles.

The academic-view dataset was constructed using the Scopus database for the purpose of uncovering key elements of academia-oriented tourism. The dataset consisted of 156,759 English language research article abstracts and titles along with their keywords, covering the period from 2000 to 2022 (precisely 30 July 2022). By analyzing the academic dataset, we identified 33 distinct parameters relating to tourism and grouped them into four main categories, viz. Tourism Types, Tourism Planning, Tourism Challenges, and Media and Technologies.

For a period of 18 months from March 2021 to August 2022, we collected a Twitter dataset showcasing the public perception of the tourism sector in Saudi Arabia. The dataset comprised of 485,813 tweets and was limited to the region to focus on local tourism issues and compare them with international views. Our analysis revealed 13 parameters, which were then categorized into 2 broader set of parameters, Tourist Attractions and Tourism Services.

A software tool (see Section 3 for details) was developed for our data-driven approach to smart tourism, composed of four software modules: Data collection and storage, pre-processing, modelling and discovery, and validation, reporting, and visualization. The tool uses pre-trained BERT word-embeddings and UMAP for dimensionality reduction, HDBSCAN for clustering, and class-based TF-IDF scores to determine the importance of words in each cluster. The parameters are discovered from clusters and grouped into macro-parameters based on domain knowledge and are validated using internal and external methods. Visualization approaches, such as dataset histograms, taxonomies, and similarity matrices, are used to describe the data and clusters. Python packages such as Seaborn, Plotly, and Matplotlib were used to construct these visuals.

Figure 1 presents a multi-perspective taxonomy of the tourism sector, generated by our software tool. This taxonomy offers both academic and public perspectives, providing a comprehensive understanding of the industry. Additionally, it includes international and national or cultural (Saudi Arabia) perspectives, highlighting the diversity of the sector. The academic and international view provides a broad and in-depth analysis, covering 15 types of tourism, various planning dimensions, major challenges, and the impact of media and technology. In contrast, the national and public perspective in Saudi Arabia focuses on services such as medical insurance and popular tourist attractions, including recent developments such as the trillion-dollar NEOM smart city, AlUla city, and seasonal festivals.

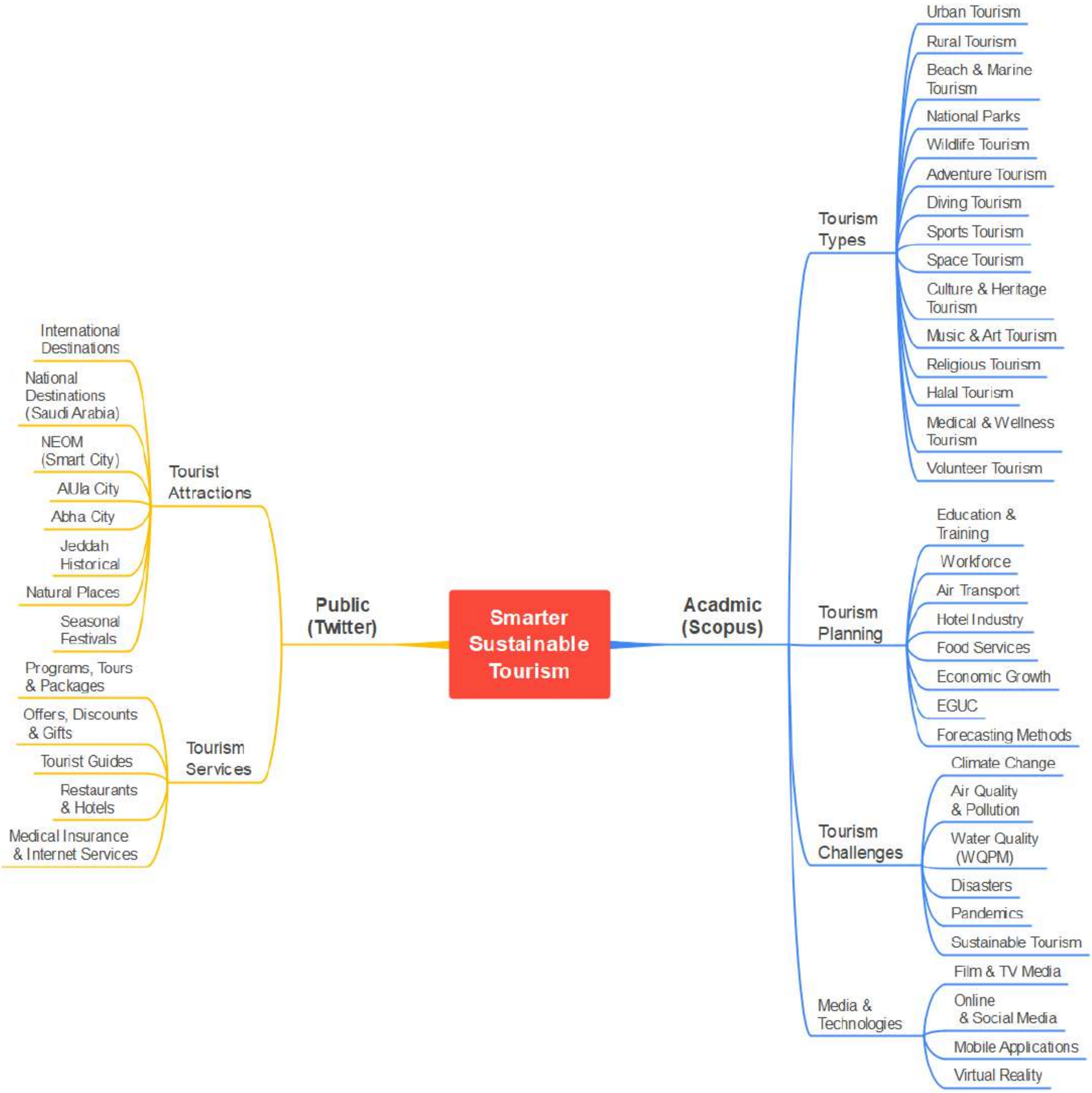


Figure 1. A multi-perspective taxonomy of smarter sustainable tourism.

The paper is organized as follows. Section 2 covers the related literature and identifies research gaps. Section 3 explains the research methodology and design of the proposed tool. Section 4 examines the parameters identified from an academic (and international) perspective using Scopus data. Section 5 examines the parameters identified from a public (and local or national) perspective using Twitter data. Section 6 presents a discussion. Section 7 concludes and suggests areas for future research.

2. State-of-the-Art and Research Gap

This section reviews works related to this paper. Our methodology that leverages deep learning has afforded us to conduct a comprehensive review of research on tourism using a dataset comprising 156,759 papers from the Scopus database. Additionally, we carried out an extensive review of research on the use of machine learning and data analytics in the tourism sector. No work directly linked to our study was found (i.e., a holistic, multi-perspective parameter discovery of the tourism sector using deep learning and big data analytics). However, to locate our study within the larger body of works on the use of machine learning and big data analytics in the tourism sector, we discuss related research from three fields. Firstly, we review research on the analysis of the academic literature (i.e., scientometric and bibliometric analysis works) in tourism. This is relevant because our work is based on the analysis of scientific literature on tourism. Subsequently, we review research in tourism on the application of artificial intelligence and machine learning for the analysis of digital and social media including Twitter. This is relevant because we analyze and extract parameters from Twitter data. Finally, we highlight the research gap in Section 2.3.

2.1. Data Analytics of Scientific Literature

Bibliometric and scientometric analysis of scientific literature has been used as an approach to analyze existing research in different areas such as finance [48], construction industry [49], transportation [50,51], smart homes [52], artificial intelligence [53,54], and others. Scientists have also employed scientometric analysis in tourism. For instance, Zach et al. [55] studied technology and innovation in the tourism sector using topic modelling. Loureiro et al. [19] used text mining to investigate augmented and virtual reality research in the tourism sector. Barrera-Barrera [56] used the LDA (Latent Dirichlet Allocation) algorithm for text mining research in tourism and hospitality with the aim to provide recommendations for selecting a suitable journal for submitting manuscripts using as criteria the manuscript topic, the journal scope, and the journal impact factor. Fang et al. [57] presented a scientometric investigation of the tourism literature with a focus on climate change using the CiteSpace tool. Chen and Zhou [58] used the CiteSpace tool and scientometric analysis to investigate people's motivation for travelling. Ribeiro et al. [59], using the VOSviewer tool, conducted a bibliometric analysis of the tourism literature with a focus on smart tourism. Baqeri et al. [60] employed scientometric analysis and the VOSviewer to investigate mining tourism with the aim of acquiring an understanding of the prospects of environmentally sustainable tourism involving the mining industry. It is clear from the literature that all the existing works have focused on specific topics in tourism. None of the works have attempted to use scientometrics to analyze the tourism literature in its broad sense and develop a comprehensive understanding of the tourism landscape.

2.2. Social Media Analytics

Data from digital and social media including Twitter has been widely used with textual analytics to study various phenomena and problems in different application domains such as in healthcare [61,62], disasters [63], smart homes [52], education [64], COVID-19-related studies [65,66], event detection [67,68], opinion mining about government services [69], city logistics [70,71], and transportation [51]. Likewise, in the tourism sector, several works involving textual analysis of Twitter data have been reported. Afzaal et al. [18] presented a classification of tourist sentiments from data in online reviews with the goal of extracting and categorizing users' favorable or negative sentiments about certain aspects. Ramanathan et al. [72] reported sentiment analysis of Twitter data with the aim of understanding tourists' views about tourism in Oman. Feizollah et al. [73] used Twitter data to investigate people's discussion topics and sentiments about halal tourism. Colladon et al. [24] reported a study based on the analysis of social networks and semantics of data from the travel forum TripAdvisor to forecast tourist arrivals in seven cities in Europe. Hasnat et al. [74] proposed a method using Twitter data for classifying tourists and residents and determining the destination preferences of tourists. Obembe et al. [75] reported sentiment analysis of Twitter data to investigate the influence of social media and communications between stakeholders on the tourism sector during COVID-19. Vecchio et al. [1] showed how value can be created from social media data to improve a tourist destination.

The works that are specific to the analysis of tweets in the Arabic language for research related to tourism are but a few. Al Sari et al. [76] used Twitter, Snapchat, and Instagram data to study the performance of different machine learning algorithms for sentiment analysis of cruise experiences in Saudi Arabia. Alasmari and Abdelhafez [77] used machine learning to analyze Twitter data to investigate visitor sentiments about tourist destinations in Saudi Arabia. Al-Smadi et al. [78] reported sentiment analysis of hotel reviews in the Arabic language. Sayed et al. [79] compared the performance of machine learning algorithms for sentiment analysis of hotel reviews in Arabic. Al Omari et al. [80] proposed a Logistic Regression method for sentiment analysis of reviews of hotels, food, and shopping places.

None of the works on social media analytics in English, Arabic, or other languages have attempted to extract a holistic national understanding of tourism.

2.3. Research Gap and Novelty

The review of the literature we provided above establishes that the current research and development on smart tourism have primarily focused on investigating specific functions, activities, or aspects of tourism, such as the analysis of tourist sentiments, experiences, satisfaction levels, cruise experiences, food and accommodation experiences, travelling routes, forecasting tourism demand, and others.

While the existing literature is rich in breadth and depth, a comprehensive understanding of the tourism landscape is missing that is needed to better design and optimize the tourism sector, particularly because of the opportunities offered by the emerging technologies that allow sector-wide and intersystem, holistic optimizations. Moreover, none of the works on Twitter data analytics in English, Arabic, or other languages have analyzed tourism in its broad sense to develop a holistic national understanding of the tourism space.

Our work bridges these gaps and is novel due to its aim, design, methods, and results. None of the earlier works have been aimed at or designed to discover a multi-perspective, international, and national, holistic landscape of tourism. None of the earlier works have used a methodology and accordingly developed a tool that uses several cutting-edge deep learning and big data methods for the analysis, discovery, and visualization of information and knowledge for policymaking, design, and operations. An outcome of this work is deep learning-based discovery of information structure and taxonomy, design, and operations parameters, and a comprehensive review of the tourism sector literature.

3. Methodology and Design

In this section, the methodology and design of our proposed system are described in detail. The system architecture is displayed in Figure 2, which examines relevant topics that describe the tourism landscape, its various dimensions, its evolving nature, digital technologies and smart societies, tourist experiences, and tourism sustainability by using Scopus research articles. The software architecture is comprised of four components that will be discussed in the subsequent section. The overview of the system, including the main algorithm, is discussed in Sections 3.1–3.6, which describe the data collecting technique, data pre-processing, parameter modelling, parameter discovery and quantitative analysis, visualization, and validation, respectively.

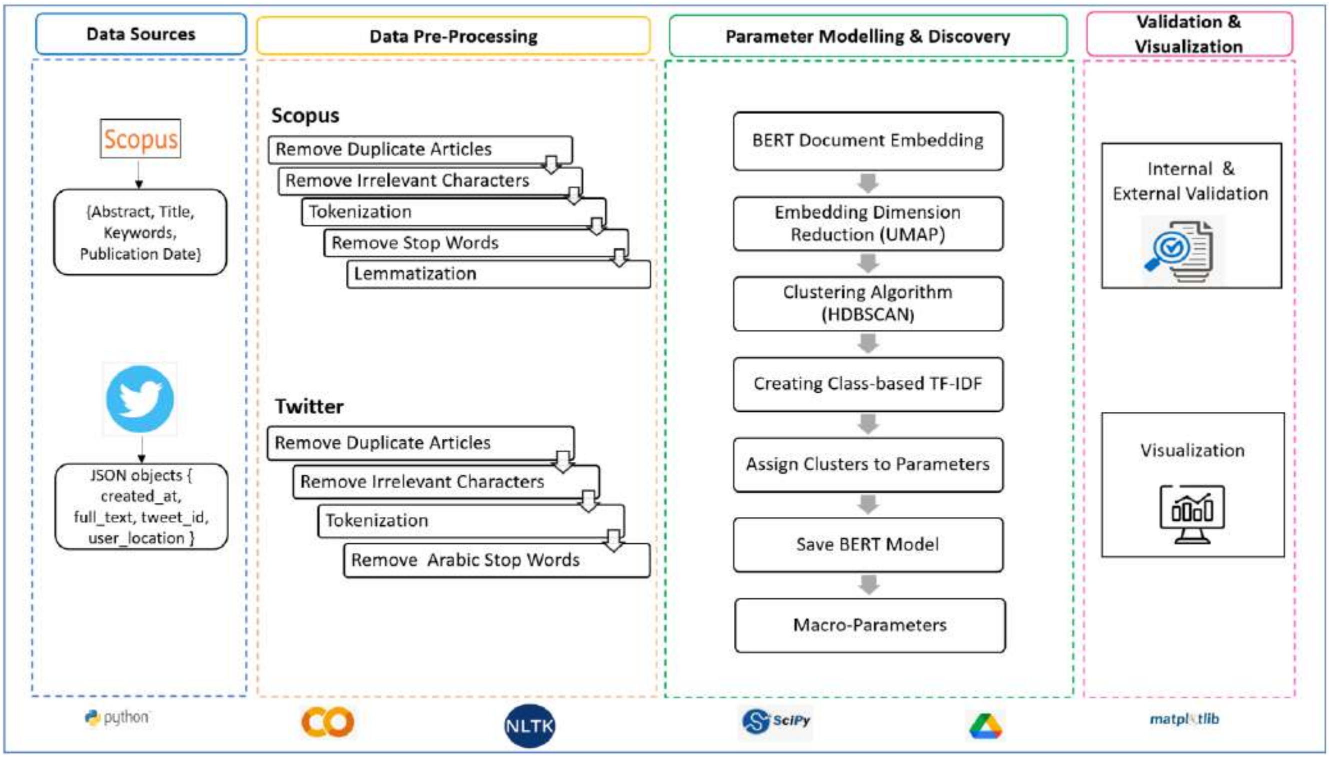


Figure 2. System architecture.

3.1. Methodology Overview

Algorithm 1 describes the fundamental steps at a high level of our proposed system. We performed a specific search query to obtain the data, then saved the results in CSV format. Then, we loaded the CSV file into the Panda data frame for the pre-processing step. During pre-processing, redundant articles, unnecessary characters, tokenization, stop words, and lemmatization are eliminated. Then, we utilized pre-trained BERT word embedding [81]. The essential part of the BERTopic model is the dimensionality reduction of the embeddings, therefore, we use UMAP. Subsequently, we applied the clustering technique using HDBSCAN to cluster the topics into groups of similar embeddings [82], and we calculate the class-based TF–IDF score to determine the significance of words within each cluster. In addition, we save the model and allocate each document to a cluster. Based on our domain expertise, we renamed the clusters as parameters and grouped these parameters into macro-parameters. The parameters and macro-parameters were thereafter displayed graphically. In addition, these parameters were validated using both external and internal validation.

Algorithm 1 BERTopic Algorithm

Input: Scopus SearchQuery
Output: Topics with labeled parameters and visualization
1: CSV_file ← DataColletion (Scopus SearchQuery)
2: AbstractPapers_DF ← read_CSV (CSV_file)
3: PreProces_AbstractPapers ← DataPreprocessing (AbstractPapers_DF)
4: word_embedding ← CreateBERT_EmbeddingModel (PreProcess_AbstractPapers)
5: UMAP_Model ← DimensionReduction (word_embedding)
6: HDBSCAN_Model ← Clustering (UMAP_Model)
7: Caclcaulate_ClassTFIDF ← Clustering (HDBSCAN_Model)
8: BERT_Clusters ← ClusterRediction (Caclcaulate_ClassTFIDF)
9: Model ← LoadModel (BERTopicModel)
10: Parameters ← relabeled (BERT_Clusters)
11: figures ← Visualization (Parameters)

3.2. Dataset Collection Method

We utilized two types of data sources. The first one is the Scopus database, which we used to collect the academic literature, whereas the second data source is Twitter. We collected the pertinent papers from Scopus, the largest database of abstracts that includes expanded data and connected academic material from many different disciplines. The knowledge provided by Scopus helps researchers, librarians, research managers, and professionals make better decisions, take better actions, and produce better results. As such, we believe it offers an academic view of smart tourism. Articles might be produced to describe public perceptions and circumstances, yet these perceptions and expressions could be called academic because they are viewed and presented by academics. We used Twitter, a microblogging social media network, to determine the public's perspective on smart tourism. Twitter might contain tweets from governments, companies, and other interested individuals. Tweets from these and other parties might be utilized to understand other perspectives. We downloaded the Scopus academic literature from the Scopus website, then we used a CSV file to save them. Moreover, we used Twitter REST API to extract Arabic tweets then stored them in a CSV file.

3.2.1. Dataset (Scopus)

We collected 156,759 research documents by utilizing the "Tourism" keyword from five disciplines: Computer Science, Social Science, Environmental Science, Engineering, Business, and Management and Accounting. We limited our research documents type to articles, reviews, conference papers, conference reviews, book chapters, and books. Furthermore, we filtered the publication years as 2000–2022 and selected the English language. Faulty data such as duplicates and "no abstract available" were removed. The final data contained 100,244 articles.

Figure 3 illustrates the histogram of the Scopus abstract articles. The x-axis displays the number of academic research articles, while the y-axis displays the word count for each research article abstract. We note that many research article abstracts contained 250 words or more. The maximum number of words in the research abstract is 1500. The average abstract length of academic research abstracts is approximately182 words. Few research articles had more than 500 words.

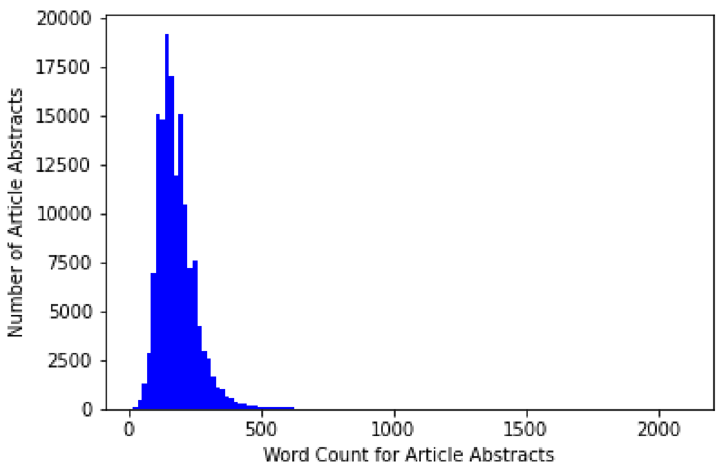


Figure 3. Histogram (Data Source: Scopus).

3.2.2. Dataset (Twitter)

The dataset collection phase used Twitter REST API to extract Arabic tweets. The Tweepy library and cursor python function were used to fetch tweets. The tweets were extracted using a search query for the keywords related to tourism, its types, and some Saudi cites, such as "السياحة" (Tourism), "السفر" (Travel), "السياحة السعودية" (Saudi Tourism), "السياحة الداخلية" (Local Tourism), "العلا" (AlUla), "نيوم" (Neom), and others. Moreover, we used several tourism types such as, "السياحةالتاريخية" (Historical Tourism), "السياحةالبيئية" (Environmental Tourism), "السياحة الدينية" (Religious Tourism), and others. Furthermore, we utilized several hashtags related to the list of tourism activities and festivals in Saudi Arabia and other hashtags for the official account of Saudi Tourism such as "#موسم_الرياض" (# Riyad Season), "#صيف_السعودية" (#Saudi Summer), and others. We also collected data using twitter official accounts that post about tourism in Saudi Arabia such as the Saudi Tourism Authority (@SaudiTourism); (@VistSaudiAR), the official account of Saudi Tourism Activates; (@GEA_SA), the official Twitter account of the general entertainment authority in Saudi Arabia; and (@Tourism_news1), the account for publishing tourism news. Table 1 indicates a sample of the keywords, hashtags, and accounts that were used for scraping the tweets. The total number of collected tweets is 485,813 tweets that were retrieved during the period from March 2021 to August 2022. We saved the tweets in CSV format. The steps of the data collection process are depicted in Algorithm 2. Firstly, we utilized a search query to search for relevant tweets using the cursor Python function. We extracted the tweet-created time, id, and text, then we saved them in a CSV file. Algorithm 2 illustrates the steps of the data collection.

Algorithm 2 Data Collection Algorithm

```
Input: Search_query {}
Output: Tweets_CVS File []
1: api ← connect_Twitter_API ()
2: Tweets_DF ← collect_tweetsTourism (created_at, tweet_id, tweet_text)
3: Tweets_CSV ← Tweets_DF
```

Table 1. The tourism keywords, hashtags, and Twitter accounts used to collect the dataset.

Keywords
السياحة، السفر، السياحة السعودية، السياحة الداخلية، سياح، مسافرين Tourism, Travel, Saudi Tourism, Domestic Tourism, Tourists, Travelers
Hashtags
#روح_السعودية، #السياحة_السعودية، #صيف_السعودية، #صيفنا_على_جوك، #السياحة_الداخلية، #موسم_الرياض، #اكتشف_العلا، #العلا ،#نيوم The Spirit of Saudi Arabia, Saudi Tourism, Saudi Summer, Our Summer, Internal Tourism, Riyadh Season, Discover AlUla, AlUla, NEOM
Accounts
وزارة السياحة، أخبار السياحة، روح السعودية، الهيئة السعودية للسياحة، الهيئة العامة للترفيه @Saudi MT, @Tourism news1, @VistSaudiAR, @SaudiTourism, @GEA_SA

Figure 4 shows the histogram of Arabic tweets. The number of tweets is indicated on the x-axis, while the word count of the tweet text is indicated on the y-axis. Tweets are commonly less than 60 words long. There are relatively few tweets with more than 60 words in them. Only a small number of tweets have a maximum of 90 words, and a few tweets have less than 10 words.

3.3. Data Preprocessing

During the data analysis, data pre-processing is an essential step after the collection of data. Data pre-processing is conducted to improve the data analysis quality and accuracy. This includes using several approaches to clean the data that has been collected. Two different types of data sources were used in our research. English academic articles were included in the Scopus dataset, whereas Arabic tweets were included in the Twitter dataset. Twitter and Scopus cannot be pre-processed using the same technique because of the varied languages utilized in both datasets. The following subsections describe the data preprocessing processes for Scopus and Twitter individually.

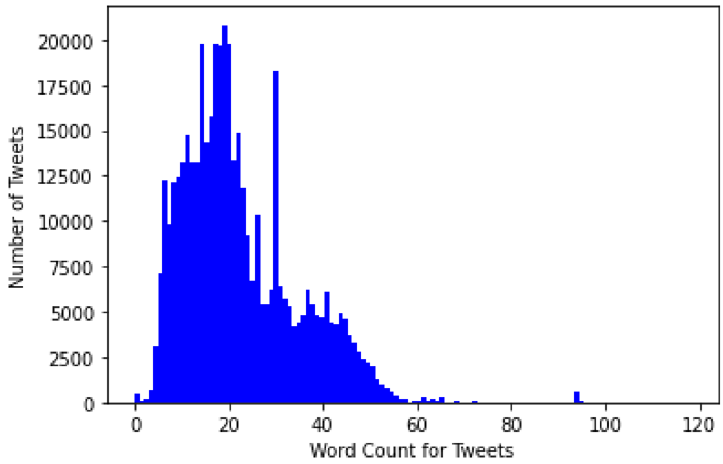


Figure 4. Histogram (data source: Twitter).

3.3.1. Pre-Processing (Scopus: English)

Among the pre-processing phases are the removal of duplicate articles, the removal of irrelevant characters, tokenization, and lemmatization using POS tags, and the removal of English stop words. Algorithm 3 illustrates the preprocessing steps as follows: First, we used a Python library (Panda) to read the CSV and saved it in the data frame. Second, we identified and removed redundant articles. Then, we deleted all unnecessary characters. In the fourth phase, the texts were tokenized using the Python package "genism", which contains the basic preprocess function. After that, the English stop words were removed using (NLTK) the Natural Language Toolkit list of predefined stop keywords. We initially developed the BERT parameter model and utilized the provided stop words list from NLTK for clustering. After obtaining the topics from the BERT model, we checked the associated keywords for each discovered parameter and investigated the irrelevant keywords that scored a high probability in the parameter. After some experimentation, we came up with a list of keywords and phrases that did not have a big impact on the parameters that were generated. Because of this, we removed certain keywords from the articles and added them to the list of stop words in our final model. Finally, the data were lemmatized using the WordNetLemmatizer. The cleaned abstracts were collected and saved in a CSV file for parameter modeling and discovery.

Algorithm 3 Data Preprocessing (Scopus: English)

Input: abstract
Output: clean abstract
1: abstract _DF ← read_CSV (CSV_file)
2: RD_DF ← DuplicateRemoval(abstract_DF)
3: RNV_DF ← RemovaNoAbstractValue (RD_DF)
4: AIC_DF ← RremoveIrrelevantCharacters (RNV_DF)
5: token_DF ← tokenizer (AIC_DF)
6: RSW_DF ← removeStopWords(token_DF)
7: lemma_DF ← lemmatization (RSW_DF)
8: clean_DF ← cleanAbstract(lemma_DF)

3.3.2. Pre-Processing (Twitter: Arabic)

Algorithm 4 illustrates the preprocessing steps of Arabic Twitter datasets. The steps are explained in the following sections.

Algorithm 4 Data Preprocessing (Twitter: Arabic)

Input: tweets_list; stopwords_List
Output: CleanTweets []
1: tweets_DF ← load(tweets_list)
2: tweets_RD ← RemoveDuplicate(tweets_DF)
3: tweets_RIC ← removeIrrelevantCharacters(tweets_RD)
4: tweetstokens ← tokenization(tweets_RIC)
5: tweets_normalize ← Normalization(tweets_tokes)
6: tweets_RS ← removeStopWords (tweets_normalize)
7: clean_Tweets [ ] ← cleantweets(tweets_RS)

Removing Duplicates Tweets and Irrelevant Characters

After collecting the tweets, we saved them in CSV files and loaded them into the Panda Data Frame format. We iterated the tweets and removed duplicate tweets to avoid the frequency of the same tweets. We defined the cleaning function that iterated the text of the tweet and removed numbers, the English alphabet, emails, URLs, extra whitespace, emojis, line breaks, and extra spaces. Moreover, all punctuations marks were removed because the majority of punctuation marks are misused. Therefore, we created a list of all punctuation containing periods (.), question marks (?), semi-colons (;), colons (:), commas (,), Arabic semicolons (؛), and Arabic question marks (?), and we removed different types of brackets, mathematical symbols, and slashes. Furthermore, we deleted the hashtags and underscores to reduce the feature set's size and increase the value of information.

Tokenization

The tokenization process involved dividing text data into words (tokens). The NLTK library contains a module called tokenize (), which further classifies sentences into sub-sentences or words.

Normalization

The normalization process involves replacing letters that have multiple shapes with the same basic shape. For example, the Arabic letter Alif (أ), which is spelled Alif, has three shapes (آ ا إ), it normalized to (ا). In additionally, the letter Yaa (ي) is normalized to (ى) dotless Yaa and Taa Marboutah (ه ة) is normalized to dot Taa (ة).

Removing Stop-Words

In any language, there are many stop-words in textual data. Pre-processing involves removing these stop-words from textual data. Certain libraries are available in several languages for text preprocessing. An example is the Natural Language Toolkit (NLTK). It is used in a wide variety of languages, including Arabic. To remove the stop-words, we used the stop-words list provided by the NLTK library. In addition, we added a list of created stop-words that are usually used in dialectical Arabic, for example, "انتو", "ايش", "عشان", and "ليش".

3.4. Parameter Modeling

We used the BERT topic modeling approach in our system to cluster the data and identify parameters [83]. We created a word-embedding model before parameter modeling by utilizing BERT, a transformer-based method created by Google [81]. BERT is used to analyze text data and extract features, such as word and sentence embeddings. The sentence-transformer model known as "distilbert-base-nli-mean-tokens" was used in our work. Our documents were initially converted into numeric representations and generated dense vector representations for each document in the data corpus. The essential part of the BERTopic model is the dimensionality reduction of the embeddings; therefore, we used UMAP, which contains a considerable number of parameters. The n_neighbors and n_components are the most important parameters. The parameter n_components refer to the dimensionality of the embedding. Depending on the circumstance, there are no direct approaches to select the optimum values. We tested the model several times, and we found that the n_neighbors = 20 and n_components = 7 produced the best result. Subsequently, we applied the clustering technique using HDBSCAN to cluster the topics into groups of related documents. Many parameters control the clustering technique, but min cluster size and min sample are the most significant HDBSCAN settings. The min cluster size specifies the minimum possible cluster size. The cluster size is controlled by the min sampling parameter. If the min sample is less than the min cluster size, the article will be merged into the same cluster. When the min sampling is large, more items are eliminated. The importance of terms for each parameter was also determined using a class-based TF-IDF score. The TF-IDF provides a way to evaluate the importance of words among texts by calculating a word's frequency in a specific document, as well as its importance throughout the whole corpus. However, if we regard each document inside a group as a distinct document prior to running TF-IDF, we will be able to evaluate the importance of each word within a cluster. This significance rating is known as the c-TF-IDF score. The parameter is more representative the more significant the words are inside a cluster. We may therefore obtain keyword-based descriptions for each parameter. Equation (1) is used to generate the c-TF-IDF score, where f is the word frequency for each class and c is determined and divided by the total number of words w. The overall frequency of words across all classes (f) is divided by the total number of unjoined texts (d) (cc).

$$\mathrm{c-TF-IDF} = \frac{f_c}{w_c} \times \log 10 \frac{d}{\sum_p^{cc} f_p} \tag{1}$$

Before the training model, it is challenging to predict how many topics will be produced once the topic model has been trained. We trained our documents using BERTopic, which produced results for several parameters. The nr_topic parameter determined the number of topics that will be reduced after training the topic model. It will reduce the number of topics to the specified nr_topics. We found the nr_topic = 45 is a reasonable number of topics. After that, we labeled all parameters and saved the model.

We re-labeled and aggregated the parameter—which was initially represented as an integer number—into macro-parameters using our domain expertise and quantitative analysis techniques. In the section below, we explain the quantitative analysis techniques.

3.5. Parameter Discovery and Quantitative Analysis

Understanding the topic model and how it works requires visualizing BERTopic and its variations. Users find it challenging to test their models since topic modeling may be a very subjective profession. An important step in solving this problem is to examine the issues and determine whether they make sense. Therefore, after our BERTopic model has been trained, we may iteratively go over hundreds of topics to truly understand the topics that were extracted. However, that takes a while and does not have a universal representation. However, visualizing the topics that were produced using quantitative analysis techniques might be a better approach. We identified the parameters and macro-parameters using domain expertise and quantitative analytic techniques, (i.e., Intertopic distance, keyword score, hierarchical clustering, and similarity matrix).

3.5.1. Term Score

A list of keywords (terms) for each parameter does not express the context of an associated parameter in the same manner. We need to determine the required number of keywords, as well as the beginning and ending points of essential keywords before we can identify a parameter. We sorted the keywords in decreasing order to visualize the c-TF-IDF score for each parameter [83]. The parameter may be identified significantly with the help of this word score visualization.

3.5.2. Inter-Topic Distance Map

The inter-Topic distance map generates a two-dimensional visualization of the parameters in an interactive Plotly graph. It is represented by parameter circles, the size of which is often related to the number of dictionary words required to describe that parameter. The circles are created using a MinMaxScaler algorithm. The parameters that are closer together share more words [83].

3.5.3. Hierarchical Clustering

The hierarchical clustering of parameters relies on cosine similarity matrices between parameter embeddings to connect parameters consistently [83]. Hierarchical clustering creates a unique cluster of hierarchical clusters by carefully matching clusters. All clusters are evaluated in all possible pairings beginning with the correlation matrix, and the pair with the greatest average inter-correlation inside the experimental cluster is selected as the new unique cluster.

3.5.4. Similarity Matrix

Applying cosine similarities to topic embeddings generates the similarity matrix [83]. The result will be a matrix displaying how closely related specific topics are to one another. The light green color displays the least similarity between parameters, while the dark blue color displays the largest similarity link.

3.6. Validation and Visualization

The outcomes may be both internally and externally verified. Internal validation of parameters entails examining and analyzing the documents associated with the parameter. Documents in our study might be academic articles or tweets. In most of the documents in our collection, we explained how we judged the relationship between the documents and the parameters. External validation is measured by analyzing the parameters, keywords, and metrics of the two datasets. We applied a variety of visualization approaches for internal and external validation. Many visualization approaches are utilized to describe the datasets, clusters of documents, and identified parameters. Dataset histograms [84], taxonomies, hierarchical clustering [85], Intertopic Distance Maps [86], similarity matrices [87], Term Rank, Similarity Matrix [88], temporal progression charts [89], and word clouds are examples of these. Several Python packages, including Seaborn [90], Plotly [91], and Matplotlib [84], are used to construct these visuals.

4. Tourism Parameter Discovery for Tourism (Academia: Scopus)

In this section, we discuss the detected parameters by our BERT model obtained from the Scopus dataset.

Section 4.1 provides an overview of the parameters and macro-parameters. In Section 4.2, we provide quantitative analysis of the clustering characteristics and discovered parameters. In Section 4.3, we discuss each individual macro-parameter in detail and the temporal analysis of each macro-parameter.

4.1. Overview and Taxonomy (Academia: Scopus)

We discovered a total of 45 clusters using the BERT modelling algorithm from the Scopus dataset. Our approach in this research was to cluster the data and then eliminate any unnecessary clusters. We omitted four clusters from the original clustering results as they were irrelevant to the work's subject. Four clusters (0, 24, 32, and 37) contain irrelevant words and the redundant keyword "tourism" due to the use of it in the research article. We grouped the 33 discovered parameters into four macros, termed macro-parameters, based on domain knowledge, the similarity matrix, hierarchical clustering, and other quantitative methods. The methodology and process used to discover parameters and group them into macro-parameters have already been described in Section 3.5.

Table 2 list the parameters and macro-parameters of the Tourism Scopus dataset. The parameters are categorized into four macro parameters, including Tourism Types, Tourism Planning, Tourism Challenges, and Media and Technologies (Column 1). Columns 2–4 list the parameters' names, their numbers, and the percentage of the articles for each parameter in the clustering model, accordingly. The fifth column represents the top 20 keywords related to each parameter. The BERT model identified 43.28% of documents as outlier clusters, so we ignored these clusters, and the remaining 56.7% of documents are listed in the fourth column.

Table 2. Macro-parameters and parameters for tourism (data source: Scopus).

Macro	Parameters	No.	%	Keywords
Tourism Types	Urban Tourism	18	0.34%	Urban, City, Urban Tourism, Culture, Space, Heritage, Branding, Tourism, Spatial, Social, City Branding, Economic, Historic, Destination, Sustainable, City Image, Policy, Urban, Planning, Design
	Rural Tourism	1	1.31%	Rural, Rural Tourism, Farm, Tourism, Agricultural, Rural Area, Village, Community, Agritourism, Farmer, Landscape, Economic, Sustainable, Cultural, Traditional, Farming, Rural Community, Land, Product, Resource
	Beach & Marine Tourism	22	0.31%	Coastal, Marine, Blue, Ocean, Area, Sea, Tourism, Maritime, Marine Tourism, Coastal Tourism, Resource, Economic, Ecosystem, Management, Island, Economy, Coastal Area, Sustainable, Coast, Coastal Zone
		31	0.24%	Beach, Coastal, Litter, Sand, Coast, Wave, Erosion, Management, Item, Area, Sediment, Marine, Sea, Along, Shoreline, Plastic, Natural, Recreational, User, Activity
	National Parks	38	0.21%	Park, National Park, National, Visitor, Area, Protected, Management, Protected Area, Nature, Recreation, Ecotourism, Forest, Area, Cultural, Sustainable, Resource, Environmental, Community, Ecological, Landscape
		39	0.21%	Forest, Specie, Are, Land, Mountain, Plant, Bird, Vegetation, Landscape, Change, Cover, Fire, Park, Tree, Habitat, Management, Ecosystem, Population, Natural, Diversity
	Wildlife Tourism	21	0.31%	Wildlife, Animal, Hunting, Specie, Human, Wild, Bear, Visitor, Population, Park, Experience, Habitat, Viewing, Area, Management, Encounter, Activity, Community, Disease, Protected
		29	0.25%	Wildlife, Animal, Elephant, Wildlife Tourism, Zoo, Specie, Human, Wild, Population, Bear, Tourism, Park, Hunting, Habitat, Visitor, Area, Management, Viewing, National Park, Encounter
	Adventure Tourism	43	0.18%	Mountain, Adventure, Hiking, Trail, Alpine, Tourism, Area, Mountain Tourism, Landscape, Climbing, Hiker, Bike, Adventure Tourism, Destinations, Mountain Area, Biking, Activity, Management, Natural, Recreation
	Diving Tourism	13	0.41%	Diving, Reef, Coral, Oil, Marine, Spill, Sea, Coral Reef, Coastal, Specie, Water, Island, Area, Site, Impact, Ecosystem, Coast, Environmental, Damage, Pollution
	Sports Tourism	25	0.28%	Olympic, Sport, Game, Event, Olympic Game, Cup, Mega, Olympics, World Cup, Host, City, Legacy, Hosting, World, Host City, Sporting, Impact, Sport Tourism, Resident, Host
		33	0.25%	Olympic, Sport, Game, Event, Olympic Game, Cup, Mega, Olympics, World Cup, Host, City, Legacy, Hosting, World, Host City, Sporting, Impact, Sport Tourism, Resident, Host
	Space Tourism	12	0.49%	Space, Flight, Launch, Commercial, Vehicle, Mission, Earth, Exploration, Station, Satellite, Private, International, Human, Technology, Cost, Design, Passenger, Law, Operation, Future

Table 2. *Cont.*

Macro	Parameters	No.	%	Keywords
	Culture & Heritage Tourism	30	0.25%	Geological, Geotourism, Geopark, Geological Heritage, Heritage, Area, Site, Volcanic, Geomorphological, Unesco, Rock, Village, Building, Historical, Economic, Architectural, Resource, Global, Mining, Activity
	Music & Art Tourism	16	0.37%	Festivals, Event, Music, Cultural, Visitor, Art, Community, Impact, Identity, Place, Organize, Culture, Economic, Group, Experience, Finding, City, Social, Held, Satisfaction
	Religious Tourism	19	0.33%	Religious, Pilgrimage, Religious Tourism, Spiritual, Pilgrim, Religion, Sacred, Church, Tourism, Christian, Secular, Spiritual Tourism, Buddhist, Spirituality, Temple, Pilgrimage, Experience, Visitor, Heritage, Place
	Halal Tourism	40	0.21%	Muslim, Islamic, Halal, Halal Tourism, Religious, Islam, Tourism, Destination, Compliant, Friendly, Country, Satisfaction, Religiosity, Non, Religious Tourism, Hotel, Attribute, Finding, Service, Saudi, Mosque, Pilgrimage, Pilgrim, Relationship
	Medical & Wellness Tourism	6	1.03%	Medical, Medical Tourism, Health, Patient, Healthcare, Hospital, Wellness, Spa, Tourism, Health Tourism, Treatment, Service, Healthcare, Travel, Quality, Country, Eastern, Industry, Medicine, Factor
	Volunteer Tourism	23	0.30%	Volunteer, Experience, Host, Organization, Community, Host Community, Participant, Motivation, Organization, Group, Project, International, Practice, Interview, Personal, Cultural, Global, Understand, Program, Motivation
Tourism Planning	Education & Training	2	1.10%	Student, Tourism, Teacher, Work, University, Industry, Management, Tourism Student, International, Language, Graduate, Classroom, Educator, Program, Vocational, Degree, Training, Experience, Questionnaire, Degree
	Workforce	17	0.34%	Employee, Job, Customer, Satisfaction, Work, Service, Organizational, Relationship, Performance, Workplace, Industry, Worker, Leadership, Organization, Manager, Engagement, Customer Satisfaction, Influence, Tourism, Working
	Air Transport	36	0.23%	Airline, Airport, Air, Low Cost, Aviation, Passenger, Carrier, Cost, Transport, Low, Flight, Service, Travel, Industry, Demand, Route, Market, International, Traffic, Code
	Hotel Industry	5	0.86%	Hotel, Customer, Hotel Industry, Industry, Employee, Performance, Guest, Manager, Satisfaction, Star, Management, Room, Green, Relationship, Hotel Manager, Star Hotel, Model, Organizational, Service Quality, Environmental
	Food Services	3	1.1%	Food, Restaurant, Culinary, Food Tourism, Cuisine, Gastronomic, Gastronomy, Tourism, Experience, Product, Dining, Consumption, Culture, Customer, International, Safety, Service, Satisfaction, Industry, Quality
	Economic Growth	15	0.38%	Growth, Economic Growth, Economic, Tourism, Model, Test, Demand, Country, Tourism Demand, International, Income, GDP, Arrival, Relationship, Result, Long, Long Run, Panel, Domestic, China
		10	0.56%	Tourism Industry, Growth, World, Growing, Economic, International, Market, Sector, Tourism Industry, Country, Largest, Travel, Economy, Fastest, Fastest Growing, Destination, Global, City, Increase
	EGUC	27	0.27%	Poverty, Pro Poor, Tourism, Alleviation, Poverty Alleviation, Pro, Community, Income, Poverty Reduction, Economic, Country, Growth, Sector, Government, Benefit, Low, Development, Rural, Policy, Household
	Forecasting Methods	14	0.42%	Forecasting, Model, Algorithm, Google, Forecast, Ontology, Forecast, Time Series, Fuzzy, Recommendation, Prediction, Neural Network, Performance, Arrival, Accurate, Combination, Result, Error, Corpus, Machine

Table 2. *Cont.*

Macro	Parameters	No.	%	Keywords
Tourism Challenges	Climate Change	44	0.17%	Climate, Weather, Temperature, Thermal, Summer, Climatic, Dog, Meteorological, Outdoor, Season, Tourism, Weather Condition, Day, Warm, Comfortable, Humidity, Tourism Climate, Winter, Daily, Warm, Humidity, Comfortable, Pet, Seasonal
		20	0.32%	Ski, Winter, Climate, Resort, Climate Change, Ice, Mountain, Sport, Alpine, Area, Season, Condition, Adaption, Impact, High, Temperature, Cover, Destination, Alp
	Air Quality & Pollution	26	0.28%	Carbon, Emission, Low carbon, climate, carbon emission, tourism, Co2, Environmental, co2 emission, Consumption, Climate, Change, Growth, Policy, Country, Travel, Economic, Greenhouse, Industry, Air
	WQPM	4	0.92%	Water, Lake, River, Water Quality, Groundwater, Water Resource, Quality, Waste, Water, Basin, Pollution, Drought, Irrigation, Management, Water Supply, Reservoir, Ecosystem, Sediment, Damage, High, Environmental
	Disasters	35	0.23%	Tsunami, Disaster, Earthquake, Damage, Recovery, Flood, Area, Landslide, Natural Disaster, Hazard, Coastal, Natural, Affected, Tourism, Vulnerability, Indian Ocean, Resilience, Building, Impact, Affected
		28	0.26%	Death, War, Site, Memorial, Literary, Tourism, Heritage, Museum, Memory, History, Visitor, Place, World War, World, Visit, Military, Book, Experience, Narrative, Right
	Pandemics	9	0.60%	COVID-19, 19, Pandemic, 19 Pandemic, Crisis, Disease, Industry, Health, Travel, Impact, Outbreak, Risk, Sector, 2020, Affected, Business Measure Global, Countries, Policy
	Sustainable Tourism	42	0.18%	Sustainable, Sustainable Tourism, Tourism, Community, Environmental, Sustainability, Amazon, Manager, Heritage, Cultural, Policy, Impact, Attitude, Destination, Economic, Protection, Visitor, Ecotourism, Industry, Management
Media & Technologies	Film & TV Media	41	0.20%	Film, Film Tourism, Movie, Television, Induced Tourism, Tv, Drama, Tourism, Popular, Screen, Video, Production, Cinema, Celebrity, Film Television, Audience, Fan, Cultural, Destination Image, Travel
	Online & Social Media	11	0.56%	Online, Internet, Website, Travel, Social Media, Golf, Facebook, User, Hotel, Customer, Marketing, Consumer, Twitter, Review, Online Travel, Booking, Instagram, Travel Agency, Sentiment, Purchase
	Mobile Applications	8	0.66%	Mobile, Smart, Smart Tourism, Smartphone, Mobile Device, Phone, City, Tourism, Technology, Mobile Application, Information, User, Smartphones, Location, Context, Experience, Map, GPS, Platform, Design
	Virtual Reality	34	0.23%	3d, Algorithm, Detection, Building, Virtual, Computer, Digital, Simulation, Software, Modelling, Sensor, Machine, Optimization, Laser, Visualization, Reconstruction, High, City, Network, Photogrammetry
		7	0.67%	Destination, Intention, Satisfaction, Vr, Image, Model, Relationship, Perceived, Effect, Service, Result, Finding, Destination Image, Influence, Customer, Quality, Factor, Emotion, Decision, Implication, Implication

The taxonomy was extracted for the 33 parameters that were detected by our BERT model (See Figure 5). The taxonomy was created from Table 2, and it indicates the parameters, their macro-parameters, and some keywords related to the parameters. The first-level branches indicate the macro parameters, the second-level branches indicate the discovered parameters, and the third-level branches show the most representative keywords associated with each parameter.

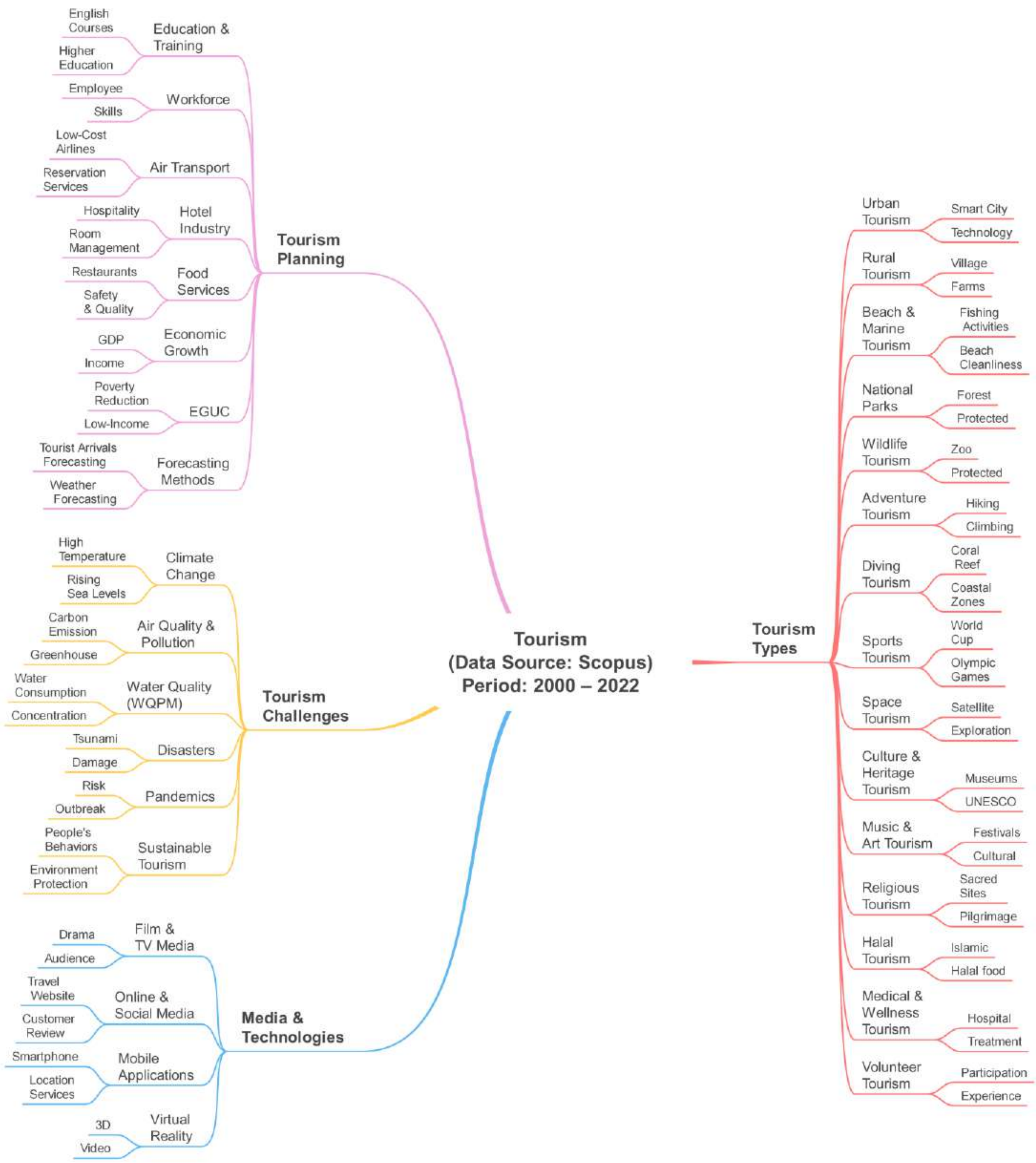


Figure 5. A taxonomy of tourism extracted from the Scopus dataset.

4.2. Quantitative Analysis (Scopus)

This section presents the quantitative analysis including the term score, word score, Intertopic distance map, hierarchical clustering, and similarity matrix. A set of keywords are used to represent each parameter; however, not all of them accurately define it. Figure 6 demonstrates that the first ten to thirteen terms in each topic's ranking accurately reflect the topic. Since the probability of all other words are so close to one another, their ordering is essentially meaningless. Therefore, in order to label the parameter, we focused on its top ten to thirteen terms.

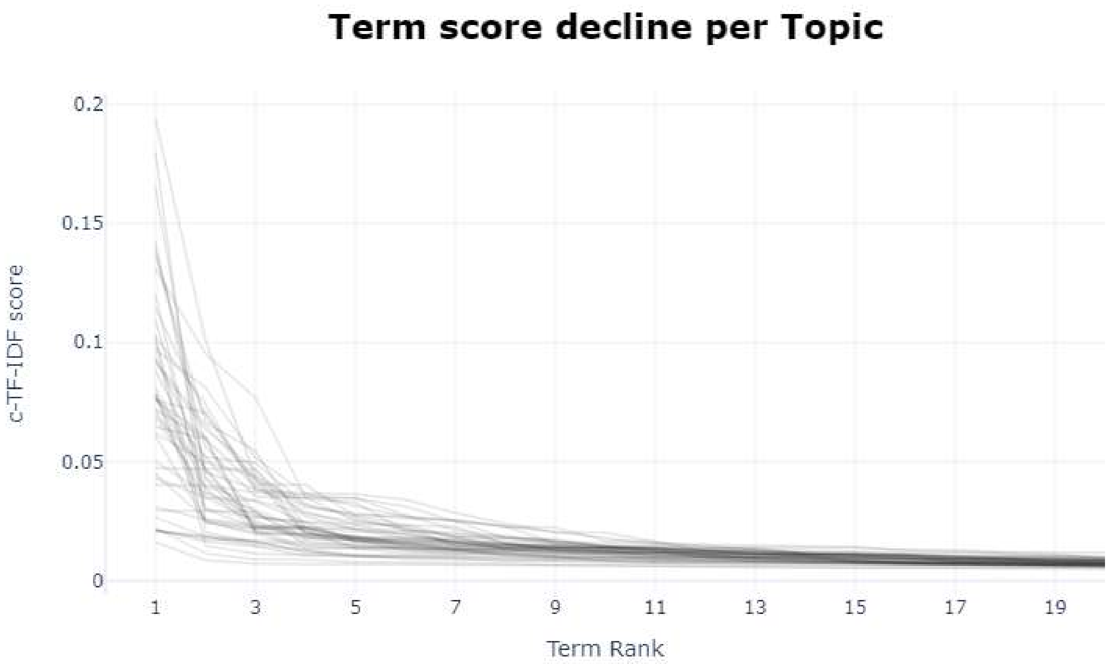


Figure 6. Term rank (data source: Scopus).

Figures 7 and 8 visualize the top 10 keywords for each parameter (see Section 3.5). The importance score, or c-TF-IDF, is used to order the keywords. There are 33 subfigures, and in each subfigure, the horizontal line shows the importance score, and the vertical line shows the parameter keywords.

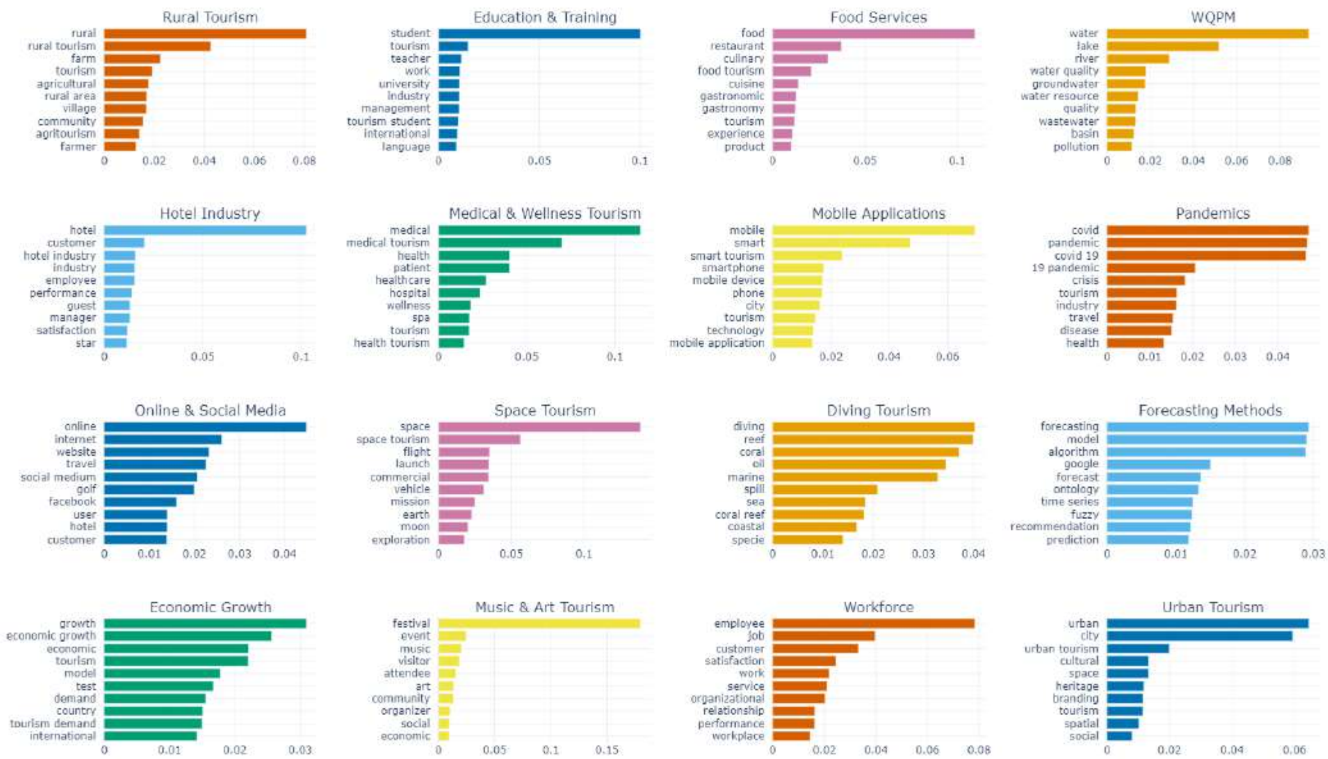


Figure 7. Smart tourism parameters with keywords' c-TF-IDF scores (data source: Scopus) (Part A).

Figure 8. Smart tourism parameters with keywords' c-TF-IDF scores (data source: Scopus) (Part B).

Figure 9 indicates the hierarchical clustering of the 33 parameters and systematically pairs them based on the cosine similarity matrix (see Section 3.5). We noticed that clusters 1, 42, 18, 43, 38, and 29 created a unique cluster that we labelled the Tourism Types parameter.

Figure 10 shows the similarity matrix among the parameters (see Section 3.5). The dark blue shows the highest similarity between parameters, whereas the light green color shows the least similarity. For example, we note there is a dark blue color between cluster 20 (Climate Change) and cluster 43 (Adventure Tourism), which indicates a high similarity score because climate change is one of the most important factors that could affect adventure tourism.

Figure 11 shows the Intertopic Distance Map based on a multidimensional scale. The figure clearly identified four macro-clusters, where three clusters are clearly identified on the right side, and the lower-left side represents one cluster.

4.3. Tourism Types

This macro-parameter captures various dimensions of the tourism landscape in relation to the following parameters: Urban Tourism, Rural Tourism, Beach and Marine Tourism, Natural Parks, Wildlife Tourism, Adventure Tourism, Diving Tourism, Sports Tourism, Space Tourism, Cultural and Heritage Tourism, Music and Art Tourism, Religious Tourism, Halal Tourism, Medical and Wellness Tourism, and Volunteer Tourism.

4.3.1. Urban Tourism

This parameter captures tourism in urban areas and its link with technologies and smart cities. The keywords in this parameter include Urban, City, Urban Tourism, Culture, Space, Heritage, Branding, Tourism, Spatial, Social, City Branding, Economic, Historic, Destination, Sustainable, City Image, Policy, Urban, Planning, Design, Creative, and Study. The dimensions and research areas related to this parameter include the mobility management of tourist flows [92], providing better tourist services [93], improving street parking systems [94], green tourism [95], strategies to develop cities with sustainable hospitality [36], and the causes and effects of over-tourism in urban cities [36].

4.3.2. Rural Tourism

The keywords that were detected by our model include Rural, Farm, Rural Area, Agricultural, Community, Area, Farmer, Village, Economic, Sustainable, Activity, Rural Community, Landscape, Resource, Social, Farming, Cultural, Business, Traditional, and Result. The Rural Tourism parameter concerns tourism involving farms, rural areas, and agricultural activities bringing many benefits to the local communities. These benefits include the economic growth of local communities [96] and sustainable development of rural areas [97]. Rural Tourism is often seen as a good option for rural development and poverty reduction [98] and a major source for creating job opportunities in rural areas [99]. For instance, farmers may benefit from tourism by offering accommodation or selling farm products to tourists [100].

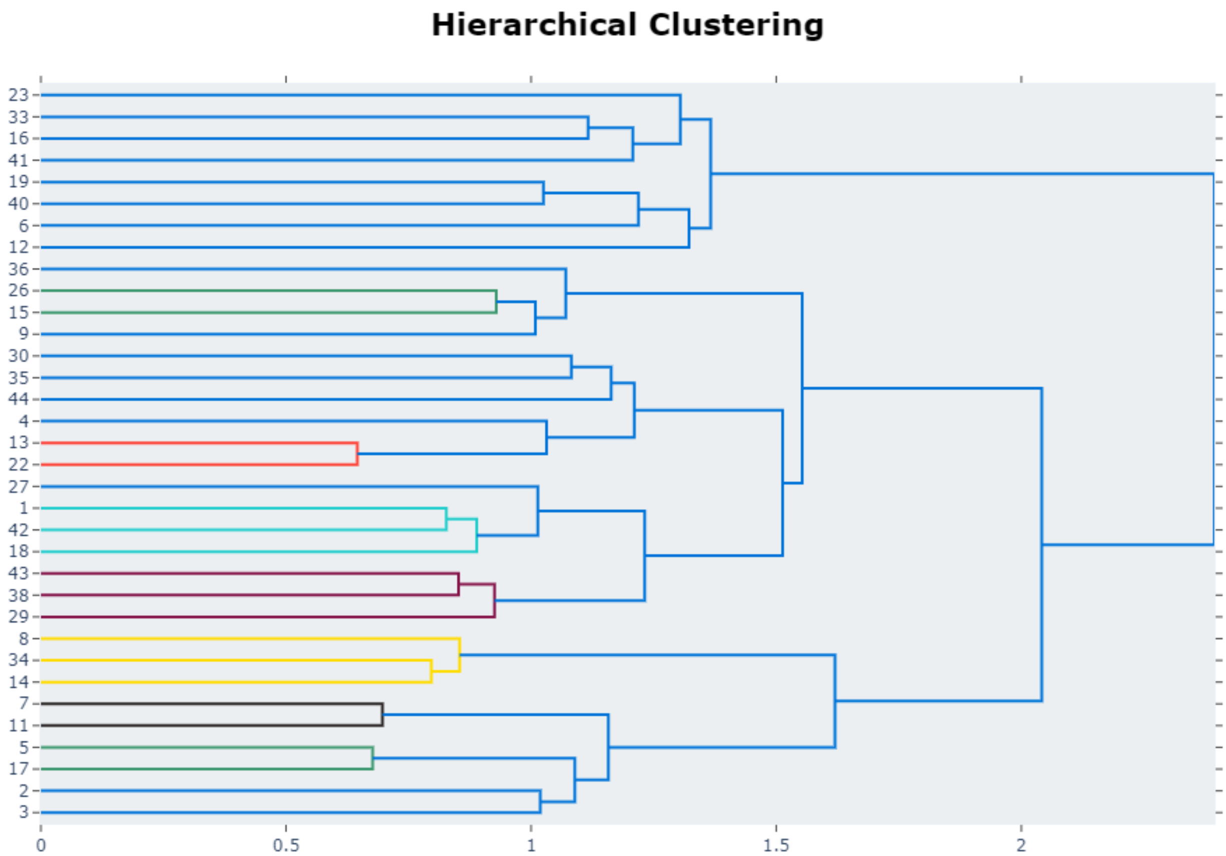


Figure 9. Hierarchical clusters (data source: Scopus).

4.3.3. Beach and Marine Tourism

The Beach and Marine Tourism parameter is created by merging two clusters (numbers 22 and 31) because the clusters include keywords pointing to similar subjects. It contains the following keywords: Marine, Island, Sea, Site, Coastal, Ecosystem, Contact, Activity, Area, Management, Recreational, Water, Specie, Impact, Damage, Ocean, Change, Cover, Environmental, and Environment. Beach and Marine are the main assets of tourist destinations. This parameter captures the issues related to the development, protection, and environmental sustainability of beaches and coastal and marine areas of the county. The parameter dimensions include developing a beach management strategy [101], ensuring the cleanliness of beach environments [102], protecting tourism infrastructure in beach zones [103], reducing environmental pollution, which can affect beach tourism experiences [104], ensuring the safety of water transportation to support the growth of marine tourism [105], and managing fishing activities in the maritime protected area [106].

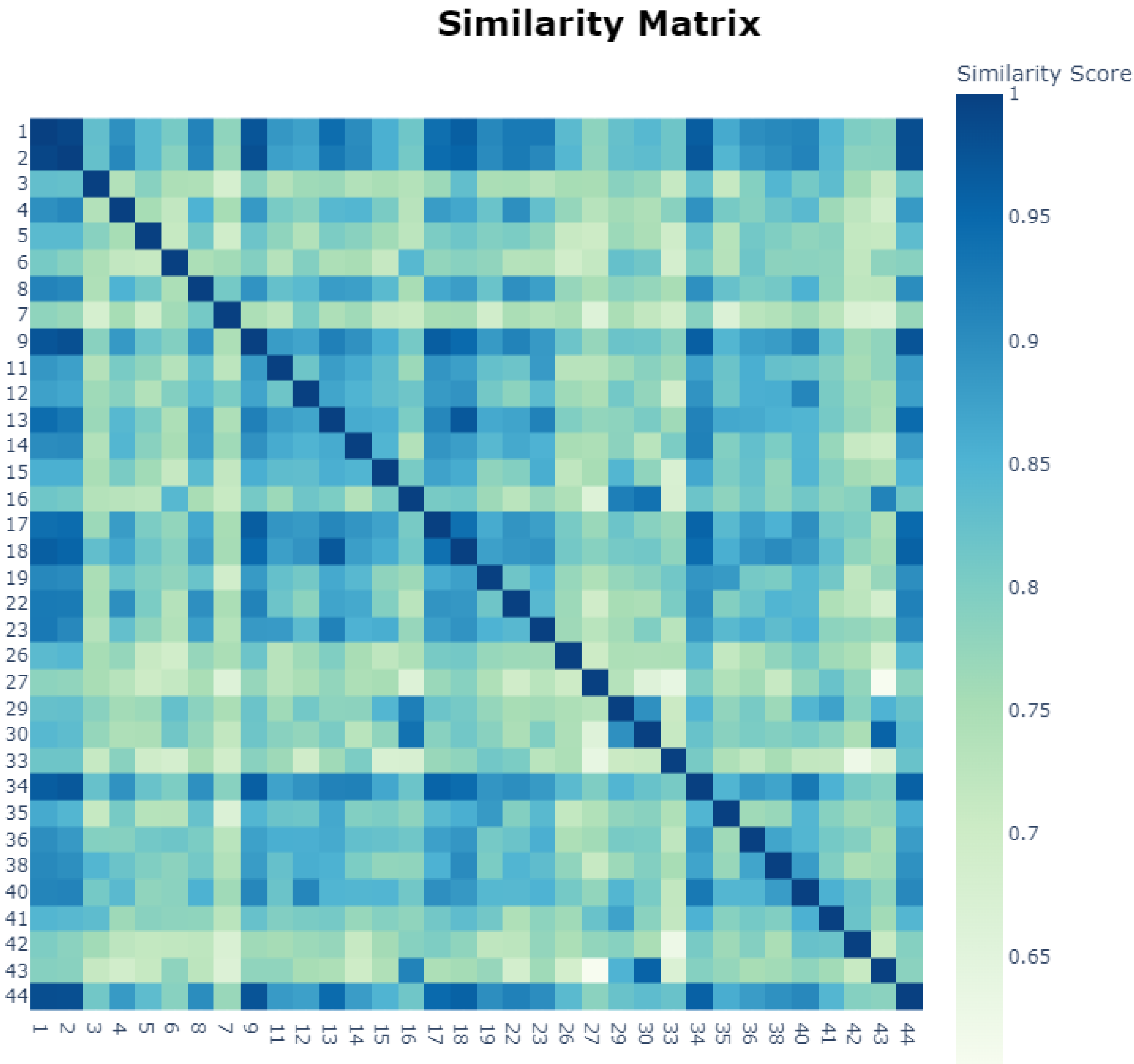


Figure 10. Similarity matrix (data source: Scopus).

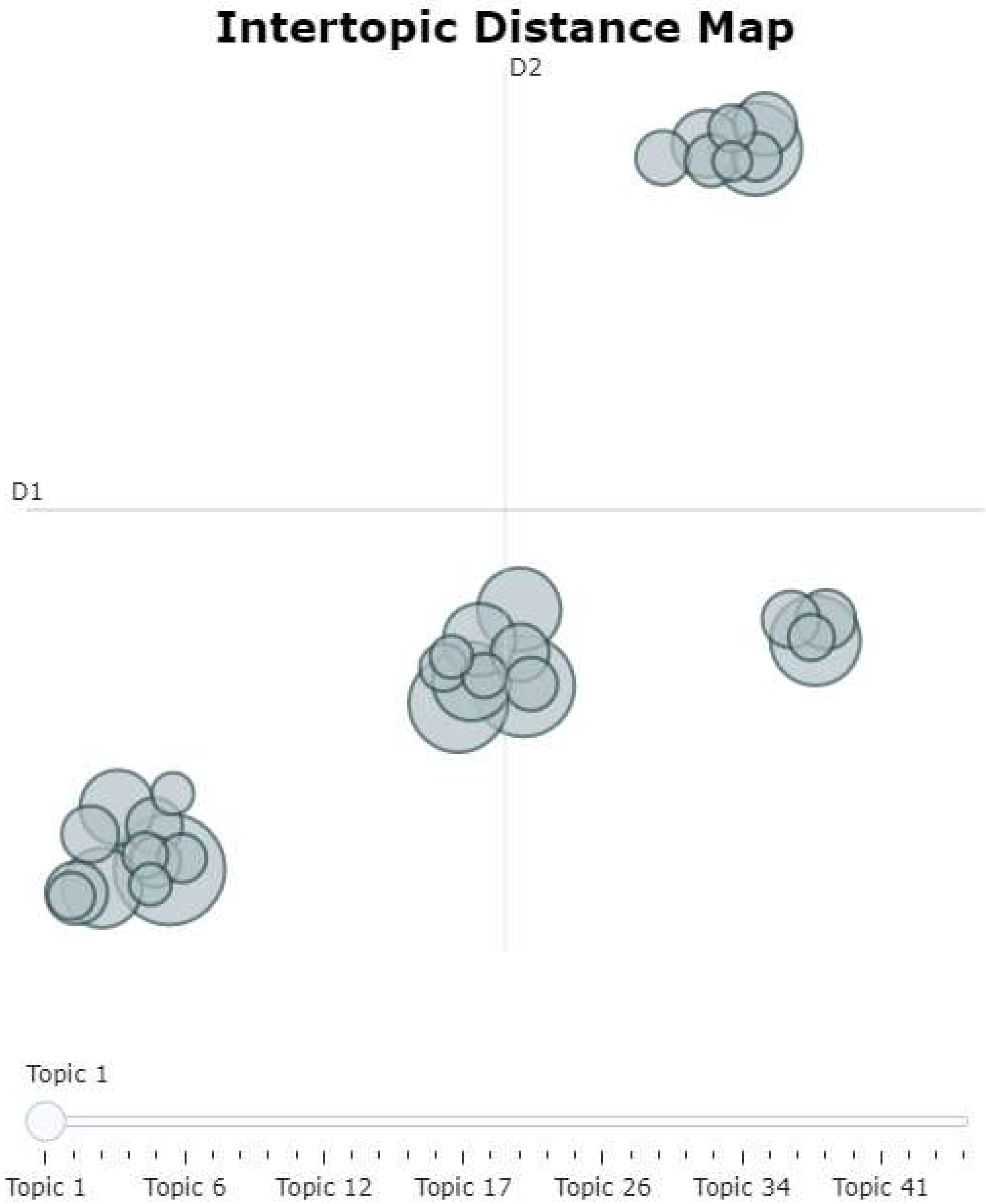


Figure 11. Intertopic distance map (data source: Scopus).

4.3.4. National Parks

The National Parks parameter explores the following keywords: Park, National Park, National, Visitor, Area, Protected, Management, Protected Area, Nature, Recreation, Ecotourism, Natural, Forest, Area, Cultural, Sustainable, Resource, Environmental, Community, Ecological, Landscape, and Village. This parameter emphasizes the importance of National Parks as tourist attractions designed to protect the natural environment from ecosystem proliferation. For example, Emphandhu et al. [107] applied a set of strategic policies to enhance local communities and National Parks. National Parks are very valuable natural areas and have the potential to attract a large number of visitors. Rogowski et al. [108] managed car traffic in National Parks to avoid over-tourism. Liu et al. [109] investigated the impact of climate change on the number of visitors in National Parks as climate change has a significant influence on rainfall and climate temperature. Kariyawasam et al. [110] investigated the role of National Parks in ensuring socioeconomic sustainability. Puustinen et al. [111] discussed how the number of visits is connected to the qualities of National Park, the leisure services within it, and tourism services in the neighboring areas.

4.3.5. Wildlife Tourism

This parameter concerns Wildlife Tourism activities in their natural habitats and tourists' experience. It is created by merging clusters 21 and 29. It contains the following keywords: Animal, Elephant, Wildlife Tourism, Zoo, Specie, Human, Population, Bear, Habitat, Visitor, Area, Management, National Park, Whale, Boat, Marine, Tour, Industry, Behavior, and Interaction. We discovered several topics related to this parameter including developing interpretations at wildlife tourist destinations such as zoos, aquariums, and others to increase visitors' understanding and appreciation for natural resources, as well as to make them aware of the effects that human activity has on animal populations [112], investigating the potential impacts of post-visit action tools such as printed handouts and email updates on the development of sustainable behavior following a wildlife tourism experience at a zoo [113], examining animal cloning and how it can help future tourist efforts [114], visiting and playing with animals in their natural environment such as watching whale and shark cage diving [115], and enhancing tourists' Wildlife Tourism experiences to achieve environmental sustainability by analyzing online reviews of wildlife tourism [116]. For instance, Flower et al. [117] looked at the types of tourists most likely to visit various elephant tourism locations and examined visitors' sentiments before and after their visits to try to raise public knowledge of the problems within elephant tourism sites in order to encourage positive attitude and behavior change.

4.3.6. Adventure Tourism

The Adventure Tourism parameter relates to travelling that involves a certain level of adventure and risk and may require specific skills such as climbing and physical exertion. It contains the following keywords: Mountain, Adventure, Hiking, Trail, Alpine, Tourism, Area, Mountain Tourism, Landscape, Climbing, Hiker, Bike, Adventure Tourism, Destinations, Mountain Area, Biking, Activity, Management, Natural, and Recreation. The dimensions in this parameter include creating an Adventure Tourism package, assisting in the improvement of product offerings in areas with natural resources [118], analyzing how visitors and tour guides behave in terms of sustainability during Adventure Tourism trips in natural places [119], evaluating the level of customer satisfaction of an adventure trip provider [120], risk management in Adventure Tourism by assisting tourism organizers and adventure tour guides in identifying risk and developing risk mitigation and risk reduction techniques [121], and investigating how Adventure Tourism may help the destination's tourist growth [122].

4.3.7. Diving Tourism

Diving Tourism relates to diving activities, coral reefs, and associated environmental impacts. It is represented by the following keywords: Diving, Reef, Coral, Oil, Marine, Spill, Sea, Coral Reef, Coastal, Specie, Water, Island, Area, Site, Impact, Ecosystem, Coast, Environmental, Damage, and Pollution. We discovered several dimensions and topics by reviewing academic documents that belong to this parameter including, among others, managing awareness of the significance of external risks to local diving operators [123], constructing artificial reefs to preserve coral reefs from several threats such as climate change, heavy human activity, and commercial usage, providing new tourist destinations that change diving and diving experiences [124], developing scuba diving guide features to ensure diver safety and environmental impact considerations associated with this activity [125], supporting the socioeconomic enhancement of coastal zones and islands by developing maritime infrastructure such as diving parks [126], and educating scuba divers to achieve environmental goals for coral reef protection [127].

4.3.8. Sports Tourism

Sports Tourism captures important dimensions of tourism related to sports events. It is represented by the following keywords: Sport, Event, Game, Mega, Sporting, Host, City, Legacy, Impact, Hosting, World, Destination, Economic, Participant, Image, Fan, Social, International, Resident, and Group. Studies under this parameter discussed several important factors related to Sports Tourism including how sports events enhance tourism by bringing foreign visitors to a city to encourage economic growth [128], how foreign world cup tourists generate significantly higher returns compared to foreign leisure tourists [129], the importance of cycling and racing sports that bring positive participation in upcoming sport events [130], building a recommendation system to notify users who are more likely to attend an upcoming active sporting event [131], and using smartphone technology to enhance tourists' viewing experiences and provide information about sport events and enable facility booking, buying of tickets, and more [132].

4.3.9. Space Tourism

The Space Tourism parameter is described by the following keywords: Space, Flight, Launch, Commercial, Vehicle, Mission, Earth, Exploration, Station, Satellite, Private, International, Human, Technology, Cost, Design, Passenger, Law, Operation, and Future. We discovered several topics linked to this parameter by reviewing the Scopus academic articles that related to it. For example, Aravindhan et al. [133] discussed the importance of medical examinations during space travel for successful survival and the difficult performance of human exploration throughout space flight. It consists of training, selecting spaceflight, and recovery from several side effects post-flight. Saputra et al. [134] studied the effect of open Space Tourism on regional development, which is measured by tourist satisfaction, destination image, and visitor attraction and commitment. In addition, the public increasingly views Space Tourism not just as a kind of entertainment but also a resource that provides access to alternative airspace, provides new opportunities for scientific and academic research, and creates opportunities for technological development and innovation [135]. Ganesha et al. [136] applied sentiment analysis from Twitter data to analyze the opinion of people about Space Tourism and found that the majority of people have confidence in space travel and eagerly await the day when they can fly to space, whereas few individuals worry and have negative feelings about the potentially harmful impacts of Space Tourism. Security is an importance issue for spaceflight [137]. Maintaining the sustainability dimensions in Space Tourism is another critical matter of concern [138].

4.3.10. Cultural and Heritage Tourism

Cultural and Heritage Tourism are among the major tourist attractions. This parameter covers concepts related to Heritage Sites, Historical Sites, UNESCO, Archaeological and Cultural Heritage, Museums, Visitor Experiences, Exhibitions, Art and History, and Augmented Reality. The parameter highlights the need for the protection of historical sites and the enhancement of tourism involving Cultural and Heritage. The dimensions in this parameter include the development of virtual museums based on emerging technologies [139], building virtual reality systems to develop virtual tours in historical tourist attractions [140], using augmented reality smartphone applications to promote historical tourism [141], developing digital storytelling methods based on recommendation systems to improve the experience of tourists who come into contact with artistic and cultural heritage [142], and educating the next generation about the historical places by using video games [143]. For example, for virtual museums, Garlandini et al. [144] illustrated how museums benefited from digital innovative technologies during the pandemic shutdown and attracted people efficiently in a safe and accessible way. Sapio et al. [143] proposed video games as an entertainment tool for historical places in Italy to teach users and the next generation about such places and enable them to interact with the world safety.

4.3.11. Music and Art Tourism

This parameter captures various dimensions of tourism involving music and art festivals. The keywords in this parameter include Festivals, Event, Music, Cultural, Visitor, Art, Community, Impact, Identity, Place, Organize, Culture, Economic, Group, Experience, Finding, City, Social, Held, and Satisfaction. A good number of research articles in this parameter covers the benefits and harms of tourism related to music events. For example, Carneiro et al. [145] studied the importance of festivals for generating economic impacts on the community and the need to define appropriate strategies to increase the positive impact. Almedia et al. [146] investigated the primary promotional channels for a music festival by understanding the primary marketing communication channels and the most efficient techniques to reach various groups of the public as a vital aspect of festival management. Montoro-pans et al. [147] analyzed the web search behavior of prospective music festival attendees, which indicates hidden patterns of behavior among cultural tourists who want to attend music festivals. Tan et al. [148] examined the influence of music festival-specific factors on visitors' levels of satisfaction and how this impacts people's well-being. On other hand, festivals can also have negative effects on host residents and communities. For instance, Moisescu et al. [149] investigated the negative impacts produced via over-tourism during the time of large music events, which include noise pollution, crime, damage to the natural environment, traffic and parking traffic problems, and others.

4.3.12. Religious Tourism

Religious Tourism is a type of tourism where people travel to visit a religious site. The Religious Tourism parameter captures studies related to religious tourism and sites. This parameter is represented by keywords including Religious, Pilgrimage, Spiritual, Pilgrim, Site, Religion, Sacred, Temple, Experience, Visitor, Heritage, Place, Motivation, Cultural, Muslim, Islamic, Tourism, Destination, and Satisfaction. Looking at the documents related to this parameter, we found different dimensions of this parameter, including religious tourist motivations [150,151], religious tourist experiences [152], exploring sacred sites [153], technological developments in religious heritage sites [154,155], and the role of Religious Tourism for enhancing social and economic growth within local communities [156].

4.3.13. Halal Tourism

This parameter captures dimensions related to Halal Tourism that is geared toward Muslims to provide them with halal food and other facilities that are compliant with or friendly to the Islamic faith. It includes the following keywords: Muslim, Islamic, Halal, Halal Tourism, Religious, Islam, Destination, Compliant, Friendly, Country, Satisfaction, Religiosity, Religious Tourism, Hotel, Attribute, Finding, Service, and Intention. Religious observances play an important role in the hotel selection process for Muslim travelers, and easily accessible Halal food is a preferred travel requirement for this group. Yen et al. [157] discuss the requirements of Muslim visitors for basic hotel and food services, specify the categories in which these features should be placed, and determine which areas require additional resources to attract more Muslim tourists. Yahaya et al. [158] analyzed the criteria related to observing Islamic rules, such as participating in fasts and prayers, eating halal food, self-discipline and abstinence, refraining from drinking alcohol and illegal sexual practices, and separating people by gender in spas and swimming pools. These are essential requirements for Muslim tourists and hotel visitors and may vary based on the individual Muslim practices.

Dabphet et al. [159] examined the criteria of Muslim tourists' selection of travel locations and accordingly looked into the overall Muslim tourists satisfaction. Rahman et al. [160] investigated the possible effects of non-Muslim visitors' perceptions of halal travel destinations and their word-of-mouth for halal travel destinations. Aji et al. [161] studied Muslims' opinions and plans to travel to non-Islamic nations and determined what variables affect Muslims' decision to travel to non-Islamic countries by examining their perceptions of those countries. Sthapit et al. [162] investigated the overall halal food preferences of non-Muslim travelers, the reasons they previously tried halal cuisines, the pleasant and negative feelings they had, and the memorable aspects of their most recent halal food experiences after returning from vacations.

4.3.14. Medical and Wellness Tourism

This parameter captures various dimensions related to tourism that is undertaken nationally or internationally primarily for the purpose of improving wellness and obtaining better or relatively inexpensive treatment of illnesses and healthcare services. This parameter is represented by the following keywords: Medical, Patient, Healthcare, Spa, Hospital, Treatment, Wellness, Service, Healthcare, Travel, Cell, Quality, Factor, Country, Eastern, Industry, and Medicine. We discovered various topics linked to this parameter by analyzing various quantitative data and reviewing the articles related to this parameter. For example, Jangra et al. [163] presented an Internet of Things (IoT)-based framework for health monitoring that may be useful for both medical travelers and hotel management in keeping track of the health of personnel and visitors. It examines numerous bodily vital signals and then informs the admin of each person's health situation. Cham et al. [164] examined how advertisements and social media communications affect hospital branding and trust building before the consumption stage, as well as their effects on medical tourists' perceptions and attitudes about service quality, satisfaction, and behavioral intentions. Mahmud et al. [165] investigated how satisfied and devoted international medical tourists are with the level of healthcare services offered by foreign medical facilities. Prajitmutita et al. [166] identified the most efficient ways for healthcare providers to allocate resources to enhance healthcare quality from the viewpoint of medical tourists. Marković et al. [167] focused on measuring customer satisfaction and service quality, particularly when entering the health tourism industry. It offers guidelines for hospital administrators to create plans that would satisfy patients' demands for high-quality care and make them more competitive in the market for Medical Tourism.

4.3.15. Volunteer Tourism

Volunteer Tourism involves volunteering to travel and benefit typically disadvantaged host communities to tackle issues such as health, education, the environment, and the economy [168]. This parameter captures various dimensions of Volunteer Tourism including volunteer tourists' participation, experiences, and motivations during their travel. The keywords represented in this parameter include Volunteer, Experience, Organization, Community, Host Community, Participant, Motivation, Group, Project, International, Practice, Interview, Personal, Cultural, Global, and Program. Examples of studies under this parameter include the participation of women volunteers in tourism in rural areas [169], examining the reasons for volunteer tourists' motivation to participate [170], how volunteer tourism organizations could encourage re-participation motivation in a short-term or long-term period [171], creating volunteering opportunities for visitors and local volunteers engaging in community festivities [172], and how volunteering tourists gain positive benefits related to their social selves and careers [173].

4.3.16. Temporal Progression (Tourism Types)

Figure 12 shows the temporal evolution of research intensity for the macro-parameter Tourism Types. The plot shows that there is a somewhat sustained increase in the research activity among all parameters (consider the y-axis scale). The parameter Rural Tourism shows the largest increase in research activity followed by Medical and Wellness Tourism. Obviously, COVID-19 has affected the growth of tourism from 2020; however, its relationship with the decline in the number of research articles needs to be established. Furthermore, we note that the decline in the numbers towards the end of the plots is because the data for 2022 are not for the whole year (the data containing research articles were collected on 30 July 2022, which makes up 58.3% of the annual data). Moreover, another interesting trend is that Rural Tourism shows a higher rate of increase than other types of tourism, and the rate is even higher around COVID-19 period, which indicates agreement with the trend because of the lower levels of social distancing requirements in open areas.

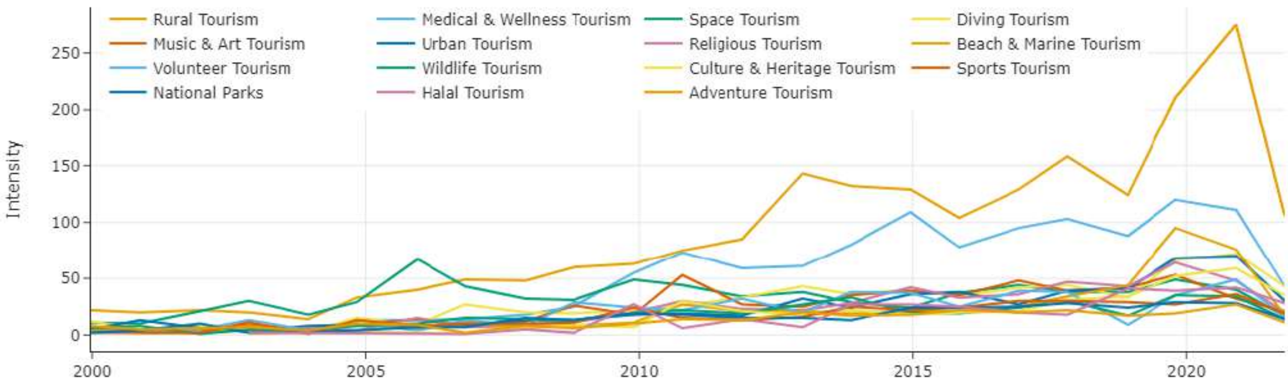


Figure 12. Temporal progression (macro-parameter: Tourism types) (data source: Scopus).

4.4. Tourism Planning

The macro-parameter Tourism Planning captures the services, planning, and development related to the Tourism Industry. It includes eight parameters, including Education and Training, Workforce, Air Transport, Hotel Industry, Food Services, Economic Growth, Economic Growth for Underdevelopment Communities, and Forecasting Methods.

4.4.1. Education and Training

This parameter relates to the development of Educational and Training programs for enhancing the tourism and hospitality industry. The keywords include Student, Learning, Teaching, Hospitality, School, Skill, Educational, Career, Curriculum, Teacher, Academic, Undergraduate, Graduate, Experience, Group, Training, Management, Higher, Industry, and University. The research topics in this parameter include students' perceptions of jobs in the hospitality and tourism industries [174], developing English courses for tourism purposes [175], creating and delivering courses in the field of sustainable tourism development to meet particular training requirements [176], and establishing models of open education for tourist companies by training experts in designing tourism areas and leisure environments [177].

4.4.2. Workforce

This parameter covers various dimensions related to the workforce in the hospitality and tourism sector including Workforce required skills, attributes, and behaviors, support structures, satisfaction, management, performance, and productivity. It contains keywords including Employee, Job, Organizational, Hospitality, Work, Hotel, Satisfaction, Performance, Workplace, Leadership, Commitment, Relationship, Working, Training, Service, Skill, Manager, Industry, Effect, and Human Resource. For example, Haarhoff et al. [178] examined the relationship between positive travel experiences and foreign language proficiency of the labor force and discussed issues to better understand how language and tourism interact. They encouraged tourist organizations to pay attention to language challenges and suggested that employment criteria for employees in the tourism sector should also consider language skills. Bani-Melhem et al. [179] indicated the most important factor influencing employees' creative behavior is their workplace satisfaction, with colleagues' support serving as a key mediating factor. Moradi et al. [180] demonstrated how better organizational commitment and e-training will enhance the success of virtual teams working in e-tourism. They also stressed that employees' participation in training initiatives plays a big part in deciding how productive and effective employees are at work.

The staff of international hotels are the most crucial links in the service delivery chain since they interact directly with the hotel's guests. Tsai et al. [181] suggest that international hotel managers must build positive relationships with their internal workforce and be effective future leaders in a dynamic environment. Choi et al. [182] investigated ways to improve the use of smart work by looking at the personal opinions of the employees.

They also demonstrated that employees did not have a comprehensive understanding of smart work effectiveness as one of the environmental protection measures in sustainability management concepts during the COVID-19 pandemic.

4.4.3. Air Transport

This parameter captures dimensions related to Air Transport (e.g., operational and management aspects) in connection with the tourism sector. The keywords include Airline, Airport, Air, Low Cost, Aviation, Passenger, Carrier, Cost, Transport, Low, Flight, Service, Travel, Industry, Demand, Route, Market, International, Traffic, and Code. This dimension and research regarding this parameter include the enhancement of airport infrastructure to meet the increased demand of passengers [183], ensuring the accessibility of direct flights to and from a location to improve the number of tourists arriving [184], measuring airport security procedures and their effects on travelers' selection of destinations during the COVID-19 outbreak [185], evaluating international tourist satisfaction when using online reservation services [186], and analyzing how low-cost airlines set their prices and how the Internet affects this strategy where both users and businesses profit from the use of internet in the purchase and sale of airline tickets and looking for the most reasonably priced flights [187].

4.4.4. Hotel Industry

This parameter covers various dimensions related to the issues and enhancement of the Hotel Industry in the hospitality and tourism sector. It contains keywords including Hotel, Green, Customer, Hotel Industry, Hospitality, Service, Performance, Guest, Quality, Room, Management, Efficiency, Business, Star, Manager, Environmental, Practice, Result, and Factor. The dimensions captured by this paper include, among others, the use of technology to provide better hospitality services, the use of technology in compliance with safety and social distancing protocols during pandemics, and the use of social media in marketing and engaging tourists, information security, and crisis management.

For example, effectively utilizing advanced and emerging technologies in the Hotel Industry influences travelers' behavior and their positive reactions to these technologies. Çakar and Aykol [188] investigated how robotic services greatly increase the quality of service provided to travelers while also positively influencing travelers' decision to return to robotics-enabled hotels in the context of customer engagement behaviors. Furthermore, in circumstances where social distance is necessary, the employment of robots in hospitality and tourist businesses will boost the mobility of individuals seeking to travel by implementing social separation through the use of robots in services. Chen et al. [189] indicated that hotel managers have to leverage social media networks to attract potential guests, since social media networks have become key contributors to customers' decision-making process when booking and visiting a hotel. Wang et al. [190] clarified how the information security policy attributes are essential for the hotel sector.

Dealing with crisis management is essential in the hotel industry. Murad et al. [191] determined the crisis management strategies employed by five-star hotels and discussed the coping and reaction methods employed to manage these crises. Martinez et al. [192] stress that hotel companies should seek to establish a green positioning strategy that can create goods and services that have both high-value and green qualities.

4.4.5. Food Services

This parameter captures research related to Food Services in the hospitality and tourism sector. Food services are an indispensable part of the tourism sector, and many see it as two sides of the same coin [193]. It includes the following keywords: Food, Restaurant, Cuisine, Destination, Gastronomy, Product, Finding, Consumption, Hospitality, Culture, Implication, Safety, Service, Satisfaction, Intention, Industry, Result, Image, and Quality. Examples of topics discovered in this parameter include the investigation of using robots in the food and beverage industry to overcome the lack of workers [194], evaluating the quality of services and customer satisfaction in the restaurant industry in which physical architecture, restaurant design, cost of the products, and the staff's response are significant factors of customer satisfaction [195], minimizing food surplus and waste in restaurants [196], the requirements of religious visitors for basic hotel and Food Services [157], and proposals for effective tourism-related food safety planning and policy [197].

4.4.6. Economic Growth

This parameter highlights the positive contribution of the tourism sector to the economic development of a country. It is represented by the following keywords: Growth, Economic Growth, Economic, Income, Country, Causality, Run, Effect, Demand, GDP, Arrival, Relationship, Model, Result, Long, Test, Long Run, Panel, and Domestic. The parameter captures various dimensions of economic growth including sustainable tourism, visitor satisfaction, host-nation well-being, safety, and prosperity, meeting tourist demands, and more. For example, Kotlyarov et al. [198] emphasized the importance of working toward the goals of sustainable development of the tourism industry to assure a high level of visitor satisfaction and achieve a balance between the host nation, the tourists, and the environment. This can lead to providing new jobs and a comfortable and safe environment, which improves the national economics and industry. Hussein et al. [199] stressed that the development of educational tourism while taking the university quality, pricing decisions, and student demand into account will make the country a center for higher education and foster economic growth and sustainability. Al-hammadi et al. [200] investigated the relationship between Halal Tourism and economic development, which has become an internationally attractive sector as a result of increased demand not only from Muslim visitors but also from non-Muslims. They stated that due to high demand, many Muslim and non-Muslim nations have taken the initiative to innovate and diversify their tourism industries by promoting Halal tourism in order to attain the satisfaction and commitment of travelers.

The number of international tourists plays a critical role in promoting Economic Growth. Din et al. [201] suggested that the economy, the number of global natural and cultural heritage sites, ethnic diversity, and strong governance are major variables affecting international tourists' destination choices. Tourism development must be based on the "green economy", which will improve job opportunities, socioeconomic growth, the protection of natural, cultural, and architectural heritage sites, and the utilization of natural resources while ensuring the renewability and sustainability of consumption [41].

4.4.7. Economic Growth for Underdeveloped Communities (EGUC)

This parameter captures tourism dimensions related to the role of tourism in developing and uplifting economically poor communities. The captured dimensions allow governments, policy makers, system designers, and operations professionals to learn the opportunities and challenges in this area and develop better solutions. The parameter includes the following keywords: Poverty, Poor, Community, Income, Economic, Country, Growth, Sector, Reduction, Government, Benefit, Low, Development, Rural, Policy, Household, Level, Social, and Political. Tourism can play a major role in developing low-income communities. Several studies have examined the issues and relationships between tourism and low-income communities. Zeng et al. [202] examined the impact of tourism on the income of poor communities and how tourism development can benefit low-income regions. Sati [203] explained that the construction of new infrastructure including hotels and homestays creates numerous job opportunities for the local community and moves it to above-poverty levels. Cole [204] investigated how tourism assists a poor, distant community's growth, and how the village benefits from the advantages of tourism without experiencing tourism's potential drawbacks. Lu et al. [205] studied the strategies that the government took to encourage rural rejuvenation through tourism, particularly in places of extreme poverty. Holden et al. [206] investigated how an understanding of poverty in communities helps provide alternative, tourism-related livelihood opportunities.

4.4.8. Forecasting Methods

The parameter captures various Forecasting Methods, research, and practices in the tourism sector. It is represented by keywords including Forecasting, Model, Forecast, Demand, Method, Google, Prediction, Neural, Time Series, Neural Network, Time, Series, Algorithm, Network, Performance, Arrival, Accurate, Combination, Result, and Error. Extensive research has been carried out on the forecasting of tourism-related variables. Examples include tourist arrival forecasting [207], rainfall and natural disaster forecasting [208], weather and temperature forecasting [34], forecasting museum visitor behaviors [209], designing tourist infrastructure, project planning for accommodations and transportation development, and reliable tourism demand forecasts to predict the number of international tourists for government agencies [210].

4.4.9. Temporal Progression (Tourism Planning)

Figure 13 shows the temporal evolution of research intensity for macro-parameter Tourism Planning. The plots show the overall growth in all parameter activities, with the highest activities in Education and Training followed by Food Services and the Hotel Industry. Note that the decline in the numbers towards the end of the plots is because the data for 2022 are not for the whole year as we collected scientific abstract research for the years of 2000 to 2022. As mentioned earlier, COVID-19 has affected the growth of tourism from 2020; however, its relationship with the decline in the number of research articles needs to be established. Moreover, the decline in the numbers toward the end of the plots is because the data for 2022 are not for the whole year (the data contained in research articles were collected on 30 July 2022, which contributes 58.3% of the annual data).

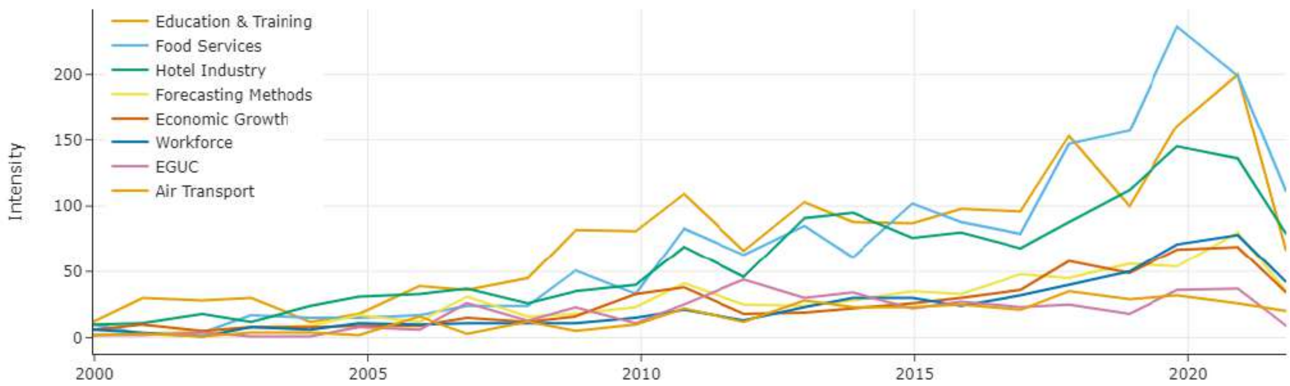


Figure 13. Temporal progression (macro-parameter: Tourism planning) (data source: Scopus).

4.5. Tourism Challenges

The macro-parameter Tourism Challenges covers major challenges facing the tourism sector. It includes six parameters: Climate Change, Air Quality and Pollution, Water Quality and Pollution Management, Disasters, Pandemics, and Sustainable Tourism.

4.5.1. Climate Change

The Climate Change parameter captures the problems and solutions surrounding the effects of the tourism and hospitality sector on Climate Change and vice versa. Climate Change and the current state of tourism have serious consequences for each other. This parameter is created from merging clusters 20 and 44. It consists of the following keywords: Climate, Weather, Winter, Resort, High, Temperature, Mountain, Season, Impact, Adaption, Risk, Destination, and Others. The dimensions and challenges covered by this parameter include extreme weather conditions [35], increasing Climate temperature [32–34], rising sea levels [211], measurement of the ozone level [212], the reduction of carbon [213], and others. For instance, an example of the extreme weather conditions dimension is the work by Younes et al. [32] who proposed a driver assistance system that recommends a safe speed during extreme weather conditions. Moreover, flight delays depend on weather conditions, and Liu et al. [214] investigated a wide range of variables that might potentially impact flight delays and suggested a gradient boosting decision tree (GBDT)-based model for generalized flight delay prediction. Hernández-Travieso et al. [34] proposed a system with a broad strategy for obtaining an accurate temperature forecast by using artificial neural networks. Gazioğlu et al. [211] examined the relationship between sea level rise (SLR), depression, and the effects of these phenomena on freshwater supplies that could be more vulnerable to the saltwater incursion. Groundwater resources are heavily utilized by the local population for urban, tourist, and agricultural water usage, which poses a serious threat to the coastal aquifer recharge.

4.5.2. Air Quality and Pollution

The Air Quality and Pollution parameter is represented by the following keywords: Carbon, Emission, Low, Energy, CO_2, Environmental, Consumption, Policy, Country, Travel, Economic, Greenhouse, Gas, and Others. The parameter relates to air pollution resulting from tourism activities where the Air Quality has a significant impact on tourism decision-making before and during the trips. International tourism is a primary source of carbon emissions [215], and traffic and public transport affect environmentally friendly development in the transportation industry [216]. Furthermore, the aircraft emissions during the cruise cycle and the landing/take-off cycle affect the Air Quality [217]. The use of green technology n the transportation sector is expected to minimize air pollution and is an active area of research.

4.5.3. Water Quality and Pollution Management (WQPM)

This parameter captures the effects of tourism on water availability, consumption, and pollution, and vice versa. It is represented by keywords including Water, Lake, River, Water Resource, Area, Quality, Resource, Basin, Pollution, Flood, Concentration, Management, Supply, Plant, Environmental, Ecosystem, Flow, and Source. Looking at the material produced by BERT including the documents that belong to this parameter we were able to find various dimensions of this parameter including sustainable water resource management [218], the impact of tourism activities on Water Quality and Water Pollution [219], water consumption in tourist pools [220], water quality prediction systems [221], and overcoming water scarcity via the treatment of wastewater and reuse of water [222].

4.5.4. Disasters

This parameter highlights how the tourism sector can be vulnerable to natural disasters and how it can solve the challenges from the Disasters that have already occurred and the risk of future Disasters. The parameter is created by merging clusters 35 and 28. It is represented by keywords including Tsunami, Disaster, Earthquake, Damage, Recovery, Flood, Area, Hazard, Coastal, Affected, Tourism, Vulnerability, Resilience, Building, Impact, Death, War, Site, Heritage, Museum, and Visitor. The dimensions and topics of the papers contained in this parameter include the impact of Disasters on tourism landmarks and the effects of frequent natural Disasters on tourist flow [223], forecasting natural Disasters [208], Disaster management recovery [224], tourist behavior after a recent Disaster in their planned destination, and how a country develops protection against potential natural Disasters [225]. For example, to protect cultural heritage from natural Disasters, Uysal et al. [226] developed a model that generates 3D photo-realistic model using photogrammetry methods for the registry. Kovačić et al. [227] investigated whether tourists' behavior is influenced by perceived risks when deciding to travel to disaster-impacted destinations. Gain et al. [224] examined the planning and recovery actions that the government has to consider during natural Disaster and crisis periods in India.

4.5.5. Pandemics

This parameter captures the impact of COVID-19 and other Pandemics on the tourism sector. It is represented by the keywords COVID-19, Pandemic, Crisis, Disease, Industry, Health, Travel, Impact, Outbreak, Risk, Sector, Affected, Business, Measure Global, Countries, and Policy. The COVID-19 pandemic has affected every element of human life, including tourism. The impression of tourism during and after a Pandemic has been altered by policies and travel restrictions established in numerous countries. The dimensions and issues related to this parameter include a sharp reduction in the number of tourist arrivals [228–230], increased travel costs and effects on the restaurant industry [231], airlines industry [232], lockdown of places of religious tourism [233], and almost entire national economies [234,235].

For example, Wang et al. [230] developed effective techniques to help predict tourist arrivals after the COVID-19 Pandemic. Ghosh et al. [229] studied how international tourism in Australia was influenced by the pandemic. They investigated the primary factors of tourism in favorable times and the principles that may be derived to re-establish international travel in the post-pandemic world. Polese et al. [231] demonstrate the importance of using technology to establish a service ecosystem supporting restaurant management. Moreover, the government set new rules and policies to overcome the COVID-19 Pandemic.

4.5.6. Sustainable Tourism

Sustainable Tourism is defined by the WTO (World Tourism Organization) as "tourism that takes full account of its current and future economic, social and environmental impacts, addressing the needs of visitors, the industry, the environment and host communities" [7]. This parameter captures various dimensions of Sustainable Tourism including Environment, Sustainability, Ecotourism, Communities, Destinations, Visitors, Residents, Homes, Impacts, Management, People's Behaviors, Norms, and Attitudes. The research activities in this parameter include, among others, managing ecotourism by protecting natural resources and environments, preserving traditional cultures of the host communities, and enhancing the environmental protection awareness of the local population [5,6], tourists' attitudes toward sustainable consumption and purchasing habits concerning eco-friendly products [8], managing over-tourism and the impacts on residents [9], and citizens' participation in environmentally sustainable actions [10]. For example, Schubert et al. [9] asserted that over-tourism has a negative impact on the cultures of the residents and consumption such as booked-out restaurants and crowded hiking trails. They stressed that maintaining investments in high-quality tourism while minimizing numbers will increase locals' well-being.

4.5.7. Temporal Progression (Tourism Challenges)

The temporal evolution of research intensity for the macro-parameter Tourism Challenges is shown in Figure 14. The plots show that there is a somewhat sustained but relatively low increase in research activity among all parameters, except Pandemics and Water Quality and Pollution Management (WQPM) (consider the y-axis scale). The relatively low growth could be attributed to the 2007 recession. The sharp increase in the Pandemics parameter is expected due to COVID-19. The relatively higher growth in the research intensity for Water Quality and Pollution Management (WQPM) reflects the significance of the problem dimension. The decline in the intensity toward the right side of the plots was explained earlier.

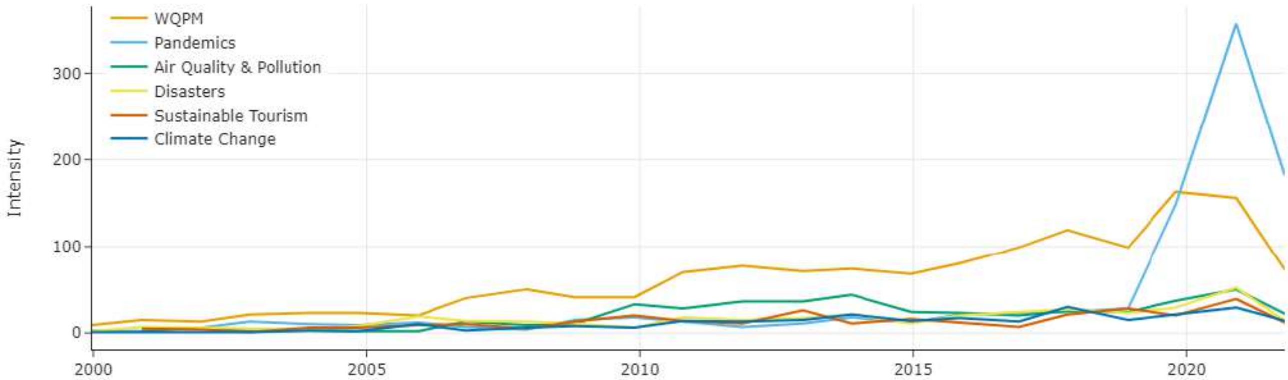


Figure 14. Temporal progression (tourism challenges) (data source: Scopus).

4.6. Media and Technologies

The Media and Technologies macro-parameter relates to various uses of Media and Technologies in development, marketing, and other functions of tourism. The use of media and technologies contributes to enhancing services and service quality, improved operational effectiveness, enhanced customer satisfaction, and cost savings. It includes the following parameters: Film and TV Media, Online and Social Media, Mobile Applications, and Virtual Reality.

4.6.1. Film and TV Media

This parameter captures the role of Film and Television Media in exploring and marketing popular tourist locations and destinations. The keywords related to this parameter include Film, Television, Induced, Destination, Image, Location, Media, Popular, Series, Viewer, Screen, Audience, Culture, Place, Phenomenon, Destination Image, Visit, and Involvement. Bolderman et al. [236] investigated the strategies that popular media produces to generate new tourism flows and emphasized how movies and music could engage the geographical imagination and physically move tourists to these locations. Nieto-Ferrando et al. [237] examined the impact of Movie-Induced Tourism (MIT) stereotypes on the satisfaction of visitors who visited a place after watching films. Iwashita et al. [238] explored the importance that films and television dramas play in influencing international visitors in selecting their travel destinations. Garrison et al. [239] studied the relationship between media tourism and sustainability and their influence on rural communities. Kay [240] used music as an instrument to promote tourism.

4.6.2. Online and Social Media

The Online and Social Media parameter captures the significant impact of Social Media applications in the tourism sector and the way it has transformed traveling and tourism. The keywords represented in this parameter include Online, Travel, Website, Social, Media, Satisfaction, Review, and Others. The tourism industry is becoming more influenced by the spread of Social Media and user-generated content. Consumers share their tourism and hospitality experiences with other consumers online by posting reviews of their stays. The articles related to this parameter disclose various dimensions of using Social Media in tourism. These dimensions include automatic text analysis and mining [76,241,242], forecasting tourist demand [24,243,244], tourist recommendation systems [245], and mining travel locations and routes [246].

4.6.3. Mobile Applications

Mobile Applications have a significant impact on tourism activities. The Mobile Applications parameter captures various developments and services that make use of mobile devices and applications. It is represented by keywords Including Mobile, Applications, Smartphone, Location, Services, Technology, and Maps. Looking at the scientific papers that belong to this parameter, we were able to find several topics that capture various mobile application services. These services include integration of the Android platform and GPS services in guiding tourists to their preferred sites [29], augmented reality for historical tourism using mobile and smart devices [30,31], the use of Mobile Applications for tracking tourist travel experiences [247], determining different types of tourists mobility and providing location-based services [20,248], and implementing QR systems for shopping centers [249].

The use of Mobile devices and Applications together with augmented reality has transformed visitor experiences by providing a chance for meaningful connection with distinctive cultures and history on smartphones. Jiang et al. [31] investigated the effectiveness of augmented reality (AR) in promoting the memorability of tourist experiences at heritage sites such as the Great Wall of China, utilizing a smartphone app. Location-based services are used as part of information systems for tour guides to provide travelers with route planning and tourism information. Yang et al. [248] developed a master multi-agent system utilizing picture recognition technology and Google Maps as a source of tourism information and a route planner for travelers. It combines a smartphone GPS function, a QR/Bar code scanner, and access to a cloud database, allowing users to locate all the necessary web services while traveling for business or pleasure. Huettermann et al. [250] discuss how visitors may use smartphone apps and obtain real-time forecasting information to resolve issues such as parking, traffic, and queues.

4.6.4. Virtual Reality

The Virtual Reality parameter is created by merging two clusters 7 and 34. The parameter is depicted by the keywords Virtual, Reality, 3D, Video, Camera, Technology, Augmented, Image, Site, Cultural, Urban, and Data. The parameter captures several dimensions represented by Virtual Reality applications in the tourism sector. These applications include web-based visualization and 3D modeling for cultural heritage objects [251], displaying a 3D image at tourist destinations [252], a virtual city tour [253], a cultural entertainment system that allows tourist to improve their travel experience by carrying out a series of games based on augmented reality [254], and virtual reality applications for museum visitors. For instance, applications were reported in [255] to visualize the grand mosque in Medina (Al-Masjid Al-Nabawi) and religious places in an interactive style with ease of use for guests with the interaction of 3D environments.

Automatic text analysis focuses on mining tourist information to improve the quality of experience for both tourists and the public tourism industry. For example, García-Pablos et al. [241] applied sentiment analysis of hotels to measure tourist satisfaction and feelings related to other factors. Al Sari et al. [76] investigated the role of cruises in Saudi Arabia using social media platforms and a machine learning algorithm. Feizollah et al. [256] applied sentiment analysis based on Twitter data, which analyzed tweets related to Halal tourism.

As regards forecasting tourist demand, for instance, Colladon et al. [24] applied social networks and predicted foreign arrivals at airports in the capital cities of Europe. Yuan et al. [257] predicted popular tourism locations and travel routes to help users obtain a better travel schedule.

This parameter also captures the impact of the COVID-19 pandemic on the hotel and lodging sector by providing a comparison of guest satisfaction before and after the pandemic [258]. The impact of COVID-19 on the airline industry was reported in [259].

4.6.5. Temporal Progression (Media and Technologies)

Figure 15 depicts the temporal progression of the macro-parameter Media and Technologies. The vertical line of the graph indicates the number of research articles, defined as the intensity. We note that the activity in the Mobile Applications parameter has significantly increased over time compared to other parameters followed by the Online and Social Media and Virtual Reality parameters. The Film and TV Media parameter has seen some growth after around 2010, although the growth is small compared to the other three parameters. These trends are in line with the expected trends and potentials of these media and technologies. Virtual reality is a high-potential technology, and its relatively low intensity is due to virtual and augmented reality applications covered under mobile applications research. The discussions about the parameters included in this macro-parameter and the plots in the figure clearly show the evolving nature of societies in relation to tourism; there is an increasing tendency of society to use mobile smart devices, social media, and virtual reality for tourism with relatively low usage of film and TV media.

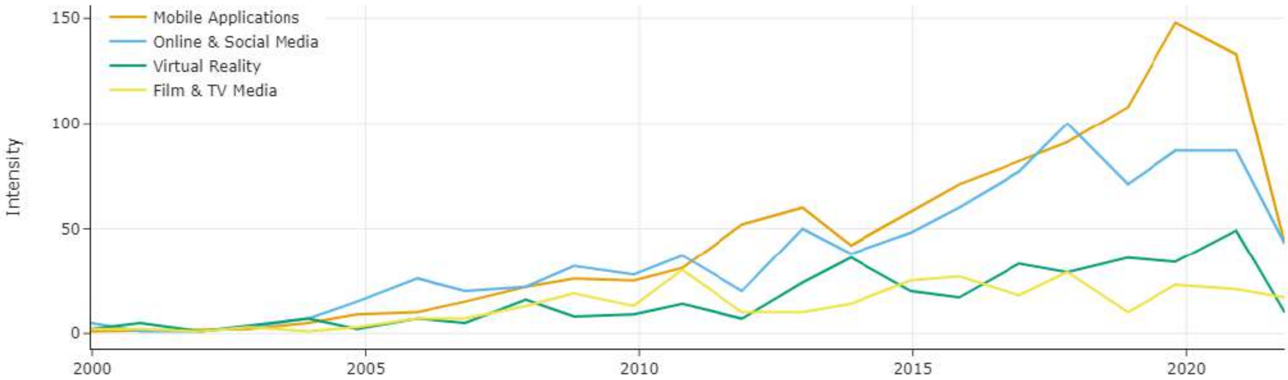


Figure 15. Temporal progression (Media and Technologies) (data Source: Scopus).

Note that some of these trends (see the lines 2019/2020 onwards) are influenced by COVID-19. The sharp drop at the end is due to the data period that does not include scientific articles for the entire year of 2022.

5. Tourism Parameter Discovery (Public: Twitter)

This section discusses the tourism discovered parameters by our BERT model from the Arabic Twitter dataset. It captures the national perspective of tourism about Saudi Arabia. The parameters are grouped into two macro parameters: Tourist Attractions and Tourism Services. We provide an overview of the parameters and taxonomy in Section 5.1. The quantitative analysis is discussed in Section 5.2. Subsequently, we discuss each macro-parameter in separate sections, namely Sections 5.3 and 5.4.

5.1. Overview and Taxonomy

We discovered a total of 30 clusters using Bert Topic modeling. Then we analyzed the results and excluded irrelevant clusters that did not have important information for our analysis. We removed six clusters due to their irrelevant themes. The remaining clusters merged as necessary and resulted in a total of 13 parameters. The parameters were grouped into two macro-parameters based on our domain knowledge, similarity matrix, hierarchical clustering, and other quantitative methods. In this section, we discuss the methodology and process used to discover parameters and group them into macro-parameters.

Table 3 lists the parameters and the macro-parameters discovered by our BERT model from the Arabic Twitter dataset. The parameters are grouped into two macro-parameters (Column 1), namely, Tourist Attractions and Tourism Services. The second and third columns indicate the parameters and their number, respectively. The fourth column lists the percentage of the number of documents. Our BERT model classified 42.7% of tweet documents in outlier clusters and we ignored these clusters, and the remaining 49.5% of tweet documents are listed in the fourth column. The fifth column lists the top 10 keywords associated with each parameter along with their English translations. As part of our efforts to gain a better understanding of the parameters, we examined the tweets associated with each parameter. As shown in the following table, we also contextually translated the Arabic tweets' content so that English readers can better understand the content.

Figure 16 shows a taxonomy of the tourism parameters that were discovered by our system. The taxonomy was created from Table 3, and it indicates the tourism parameters, their macro-parameters, and some keywords related to the parameters. The first-level branches represent the macro-parameters including Tourist Attractions and Tourism Services. The second-level branches represent the discovered parameters such as International Destinations, National Destinations (Saudi Arabia), NEOM, Programs, Tours and Packages, Restaurants and Hotels, etc. The third-level branches represent the most representative keywords associated with each parameter.

Table 3. Macro-parameters and parameters for tourism (data source: Twitter).

Macro	Parameters	No	%	Keywords
Tourist Attractions	International Destinations	1	3.04%	أذربيجان، قابلا، قوبا، باكو، بورجومي، دبي، باتومي، تبليسي، قطر، البحرين Azerbaijan, Qabil, Quba, Baku, Borjomi, Dubai, Batumi, Tbilisi, Qatar, Bahrain
	National Destinations (Saudi Arabia)	0	14.20%	السياحة، السفر، الرياض،نيوم، الداخليه، موسم، العالم، جده،المدينه، العلا Tourism, Travel, Riyadh, Neom, Local, Season, The World, Jeddah, Medina, Alula
		25	0.85%	خيبر، بدر، ينبع، العلا، المنوره، المدينه، جازان، البكيرية، للتفاصيل، استعلم Khyber, Badr, Yanbu, Alula, Munawwarah, Medina, Jazan, Bukayriah, Details, Inquire
	NEOM (Smart City)	8	1.03%	نيوم، الأكثر، الاحلام، مستوي، المشروع، طموحا، العالم، للقائد، والملهم، المستقبل NEOM, The Most, Dreams, Level, Project, Ambitious, World, Leader, Inspiring, Future
		13	0.92%	نيوم، جده، مطار، الرياض، برج، المراقبه، فرضت، مغلقه، واشنطن، المرضي NEOM, Jeddah, Airport, Riyadh, Tower, Observation, Imposed, Closed, Washington, Accepted
		22	0.66%	المستقبل، والاستثمار، نيوم، يعلن، تصاميم، اهداف، أطلقها، لاين، الاقتصاديه، فكرتها Future, Investment, NEOM, Announces, Designs, Goals, Launched, Line, Economic, Idea
	AlUla City	3	2.14%	العلا، املج، قمة، مدائن، جده، قرارات، اعلى، الخليجي، وزراء، بتبوك AlUla, Umluj, Summit, Mada'in, Jeddah, Decisions, Higher, Gulf, Ministers, Tabuk
	Abha City	14	0.96%	ابها، عاصمه، طرق، لعام، ولمنطقه، للمدينة، انجاز، يعتبر، عسير، ينبع، السياحه Abha, Capital, Roads, Year, Region, For City, Achievement, Considered, Asir, Yanbu, Tourism
	Jeddah Historical	18	0.75%	التاريخيه، جولة، جده، تستضيف، والجغرافية، بمكانتها، أهميتها، مساعده، مساعد، الجيولوجية Historical, Tour, Jeddah, Host, Geographic, Position, Importance, Assistance, Helper, Geologic
	Natural Places	26	0.55%	القوقاز، جبال، الأوربية، الطبيعة، جنه، المسافر، يلمسها، امتزاج، حلما، الخلابة، تمتع Caucasus, Mountains, European, Nature, Paradise, Traveler, Touches, Mix, Dream Wonderful
	Seasonal Festivals	11	0.99%	تذاكر، احجز، ونترلاند، موسم، جده، الرياض، البوليفارد، يجي، تتوفر، الي يبها، والبليفارد Tickets, Book, Winterland, Season, Jeddah, Riyadh, Boulevard, Come, Available, If You Want
		16	0.92%	تذكرتك، فعاليات، احجز، جنغل، جده، تذاكر، حفل، حتعيش، تتاخر، والموسيقى Your Ticket, Events, Book, Jungle, Jeddah, Tickets, Party, Live, Be Late, Music
		17	0.91%	السنه، تذاكر، راس، موسم، الرياض، حفله، البوليفارد، العلم، جزيره، ونترلاند Year, Tickets, Head, Season, Riyadh, Party, Boulevard, Flag, Island, Winterland

Table 3. *Cont.*

Macro	Parameters	No	%	Keywords
Tourism Services	Programs, Tours & Packages	5	1.49%	سياحي،سياحيه ، برامج ، سيارة ، برنامج ، انسب، حجز فنادق، تأجير،متميزة ، جولات، حجوزات Tourist (a person), Tourist (adjective), Programs, Car, Program, Most Suitable, Hotel Reservation, Rental, Distinguished, Tours, Reservations
		15	0.93%	كشتات، مرافقه ، بوكنيك، توصيل ، حجز، سائق،شقق ، دليل ، المطار، بأسعار Camping, Accompany, Booking, Delivery, Reservation, Driver, Apartments, Guide, Airport, Prices
		27	0.55%	المسافرون، رحلات، العرب، عمان، عوائل، باتومي ، عطلات، البحرين ، للحجز، قطر Travelers, Trips, Arabs, Oman, Families, Batumi, Holidays, Bahrain, For Reservations, Qatar
	Offers, Discounts & Gifts	29	0.50%	عرض، سافر، احجز، واستمتع ، ووفر ، بأسعار ، سفر ، قطر، منتجع ، جميل Offer, Travel, Book, Enjoy, Save, Prices, Travel, Qatar, Resort, Beauty
		2	2.22%	أذربيجان، خاصه ، خصومات، السياحه ، للشركات، رياده ، وهدايا، الكبيرة ، استفسار، الشتاء Azerbaijan, Special, Discounts, Tourism, Companies, Entrepreneurship, Gifts, Big, Inquiries, Winter
		21	0.70%	أفضل ، العروض ، الأسعار، الالكتروني ، والاسعار، دوله ، عشاق، اثناء، أذربيجان، الخصومات Best, Offers, Prices, Electronic, Prices, Country, Lovers, During, Azerbaijan, Discounts
	Tourist Guides	4	1.53%	والاستعلام، مرشد ، رحلتك، عروض ، للحجز، الامارات، البحرين ، عمان، تبليسي ، قطر Inquire, Guide, Your Trip, Offers, For Reservations, Emirates, Bahrain, Oman, Tbilisi, Qatar
	Restaurants and Hotels	23	0.62%	الأماكن،السياحيه ، جهزت، والمواقع ، عروس ، والمطاعم،كامل ، الأوسط ، مدن ، الخليج Places, Tourist, Prepared, Sites, Bride, Restaurants, Full, Middle, Cities, Gulf
		24	0.58%	ألفخم ، الفنادق، المطاعم ، اسكن ، السائحين، القريب، ارقي ، عرض ، دوله ، اجازتكم Most Luxurious, Hotels, Restaurants, Live, Tourists, Nearest, Finest, Offer, Country, Vacation
		7	1.03%	عروض ، خدمات ، أسعار، الشتاء، السياحه ، منتجعات،أفضل ،والأمان، الراحه ، الأماكن Offers, Services, Prices, Winter, Tourism, Resorts, Best, Safety, Comfort, Places
	Medical Insurance & Internet Services (MI2S)	12	0.99%	تامين، واحصل ، رحله ، وشرائح ، انترنت، اخبروه ، لسائق، رأيتم ، تاكسي ، نشرت Insurance, Get, Trip, SIM Cards, Internet, Tell Him, Driver, You Saw, Taxi

5.2. Quantitative Analysis (Twitter)

This section presents the quantitative analysis including the keywords score, the Intertopic distance map, hierarchical clustering, and the similarity matrix. A set of keywords are used to represent each parameter; however, not all of them accurately define it. Figure 17 presents the top 10 keywords for each parameter (see Section 3.5). The importance score, or c-TF-IDF, is used to order the keywords. There are 13 subfigures, and in each subfigure, the horizontal line shows the importance score, and the vertical line shows the parameter keywords.

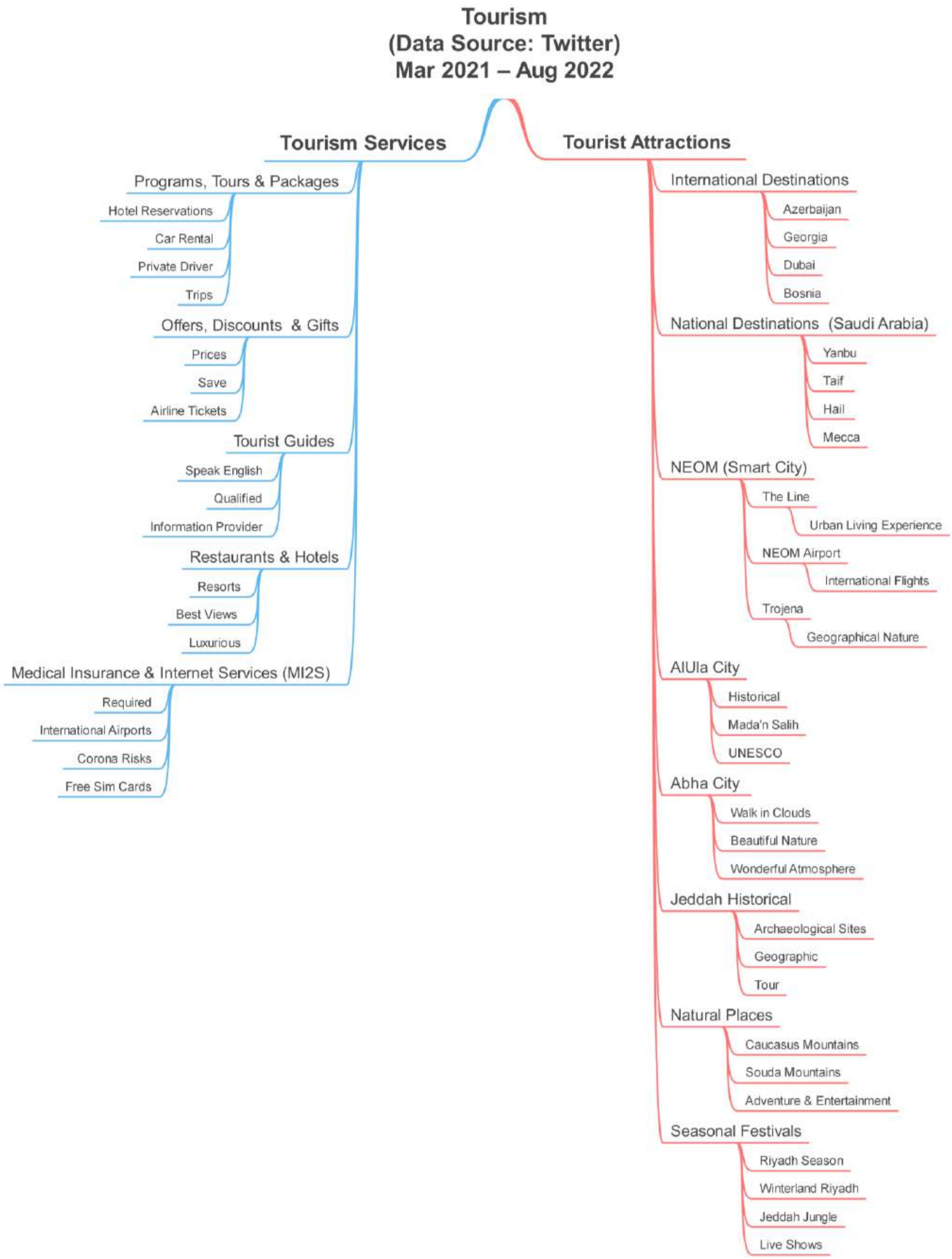


Figure 16. Taxonomy of discovered tourism parameters extracted from Twitter data.

Figure 17. Twitter parameters with keywords' c-TF-IDF score (data source: Twitter).

Figure 18 indicates the hierarchical clustering of the 13 parameters and systematically pairs them based on the cosine similarity matrix (see Section 3.5). We noticed that clusters 0, 3, 14, and 22 created a unique cluster that we labelled Tourist Attractions.

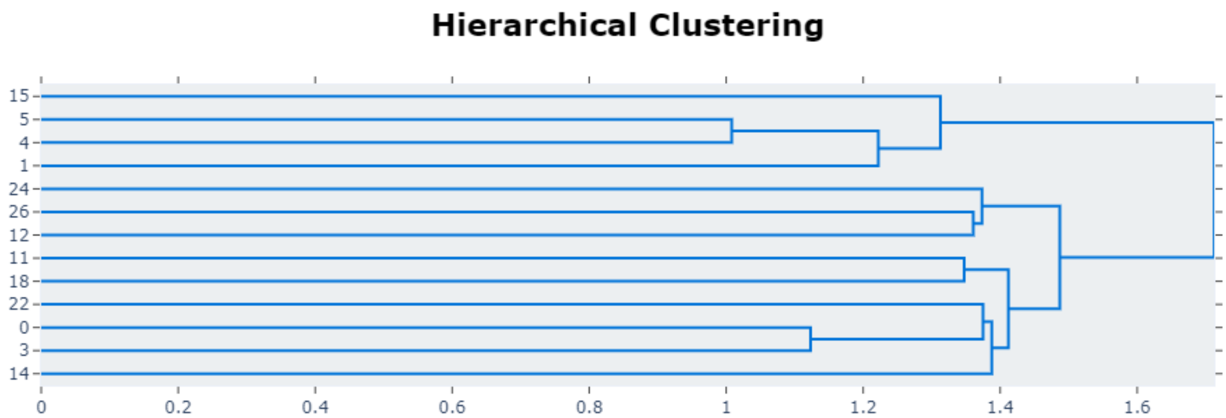


Figure 18. Hierarchical clustering (data source: Twitter).

Figure 19 shows the similarity matrix among the parameters (see Section 3.5). Dark blue shows the highest similarity between parameters, whereas light green color shows the least similarity. For example, we see a dark blue color between clusters 11 and 18, which indicates a high similarity score because both clusters 18 (Jeddah Historical) and 11 (Seasonal Festivals) are related; Jeddah city had seasonal festivals (Jeddah Season) during the period of 2 May 2022 to 30 June 2022.

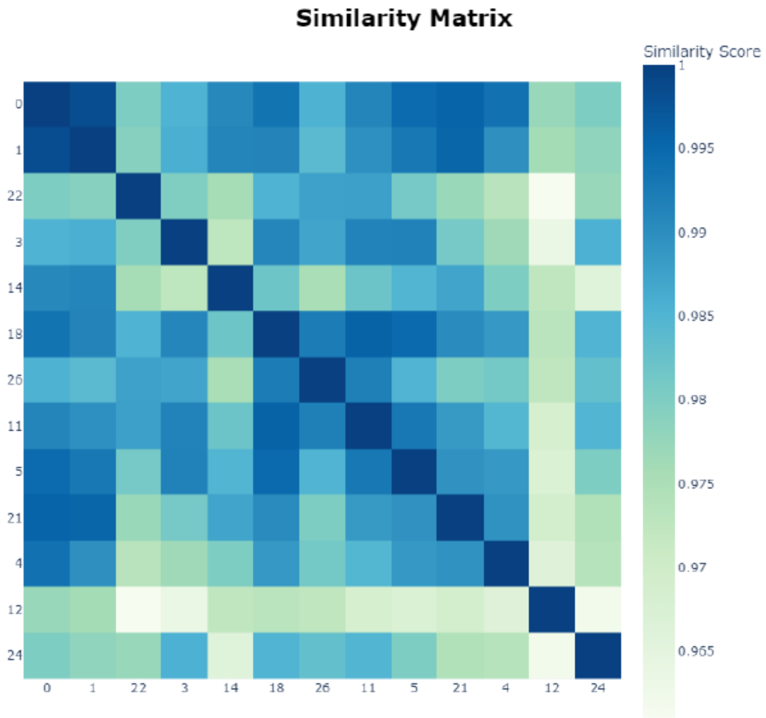


Figure 19. Similarity matrix (data source: Twitter).

Figure 20 shows the term scores that identify the number of keywords needed to describe the parameters. It indicated that ten to thirteen terms in each topic's ranking accurately reflect the topic. Since the probability of all other words are so close to one another, their ordering is essentially meaningless.

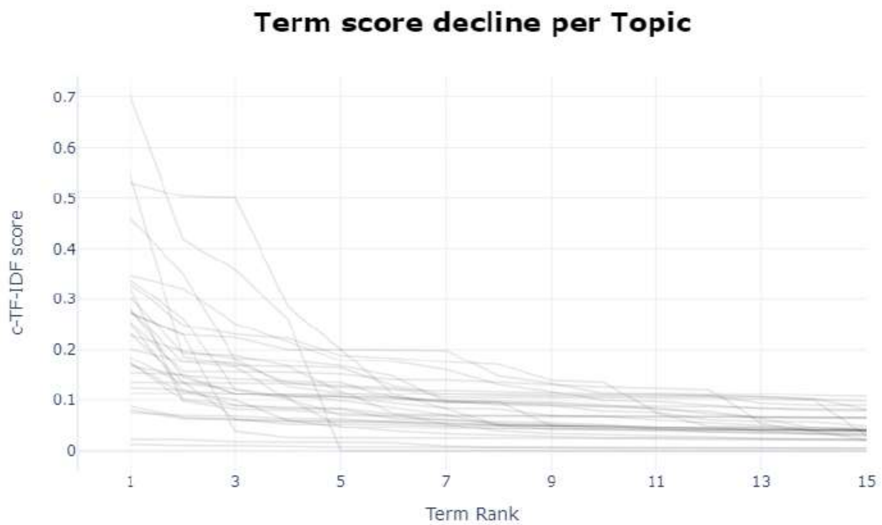


Figure 20. Term rank (data source: Twitter).

Figure 21 shows the Intertopic distance map based on a multidimensional scale. The figure clearly identified three macro-clusters.

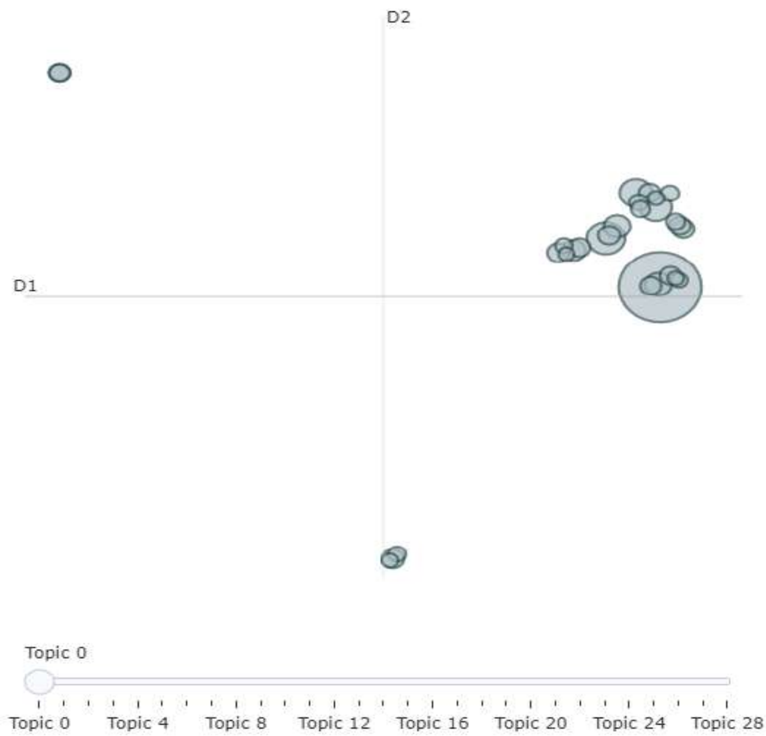


Figure 21. Intertopic distance map (data source: Twitter).

5.3. Tourist Attractions

In this section, we discuss the first macro-parameter Tourist Attractions. It includes eight parameters, namely International Destinations, National Destinations (Saudi Arabia), NEOM (Smart City), AlUla City, Abha City, Jeddah Historical, Natural Places, and Seasonal Festivals.

5.3.1. International Destinations

The first parameter is International Destinations. It is represented by keywords including Azerbaijan, Qabil, Quba, Baku, Borjomi, Dubai, Batumi, Tbilisi, Qatar, Bahrain, travel, Emirates, Tourism, Bosnia, and Trabzon. An example of a tweet related to this parameter from a company is as follows:

"للمزيد من المعلومات عن السياحة فى#اذربيجان#باكو#جورجيا#تبليسي#باتومي#قوبا#قابالا#بورجومي#سفر
#سياحة#دبي#الامارات"

"For more information about tourism in #Azerbaijan #Baku #Georgia #Tbilisi #Batumi #Quba #Qabala #Borjomi #Travel #Tourism #Dubai #UAE"

5.3.2. National Destinations (Saudi Arabia)

The second parameter is National Destinations (Saudi Arabia). It is created by merging two clusters (numbers 0 and 25). This parameter is related to popular tourism destinations in Saudi Arabia. It includes the following keywords: Tourism, Travel, Riyadh, Neom, Local, Season, Jeddah, Medina, AlUla, Mecca, Ministry, Yanbu, Tabuk, and Dammam. The following tweet mentions the hashtags of several tourism destinations in Saudi Arabia. The second tweet is from the Saudi Ministry of Tourism and it highlights the trends towards national and ecological tourism due to COVID-19 risks and physical distancing restrictions:

"#طيران_اديل#السعودية#الرياض#جدة#مكة_المكرمة#المدينة_المنورة#ابها#تبوك#حفرالباطن#الدمام#الخبر
#القصيم#بريدة#الزلفي#الجوف#المجمعة#حائل#الطائف#ينبع#نجران#الخرج#الخفجي#العلا#سياحة"

"Flyadeal #Saudi Arabia #Riyadh #Jeddah #Mecca #Abha #Tabuk #Hafar Al-Batin #Dammam #Khobar #Qassim #Buraydah #Zulfi #Jouf #Al Majmaah #Hail #Taif #Yanbu #Najran #Kharj #Khafji #Ula #Tourism"

"من الأمور الايجابية التي نتجت عن كورونا هي السياحة البيئية والدينية الداخلية التي نشطت بشكل كبير وهنا دور وزارة السياحة بالتنسيق مع النقابات ووكالات السفر والسياحة"

"Among the positive things that resulted from Corona is the internal and ecological tourism, which has been very active, and here is the role of the Ministry of Tourism in coordination with unions and travel and tourism agencies"

5.3.3. NEOM (Smart City)

The third parameter is Neom (Smart City), a trillion-dollar smart city being built in the northwest of Saudi Arabia. NEOM intends to bring about a shift in Saudi economy by developing a knowledge-based economy, diversifying the country's economic sources of revenue, and reducing the country's reliance on oil. This parameter is represented by the following keywords: NEOM, Dreams, Project, World, Inspiring, Future, Renewed, Ambitious, Inspired, and Leader:

"مشروع نيوم هو المستقبل والتاريخ الجديد وهو الوجهة الأكثر ملاءمة للعيش على مستوى العالم حيث يسعى لتنمية الاقتصاد السعودي وتنويع مصادره، وتقديم نموذج مثالي لمعيشة مستدامة ومزدهرة."

"NEOM is the new future and history, the world's most livable destination. It seeks to develop the Saudi economy, diversify its sources, and provide an ideal model for sustainable and prosperous living".

"رئيس قطاع السياحة في شركة #نيوم من المخطط افتتاح أول فنادق مشروع "نيوم" في نهاية عام 2022، كما سنفتتح ما يصل إلى 15 فندقًا سنويًا في الفترة ما بين 2023 و 2025"

"Head of the tourism sector in #NEOM: It is planned to open the first hotel in the NEOM project at the end of 2022, and open up to 15 hotels annually between 2023 and 2025"

"نيوم تتميز بطبيعة جغرافية فريدة من نوعها بالعالم كله ومن اشكال الاستغلال الأمثل لهذي الطبيعه جات مشاريع عملاقة في نيوم مثل اطلاق مشروع #تروجينا واللي راح يصير الاميز في السياحة البيئية"

"NEOM is characterized by a unique geographical nature in the whole world, and among the forms of optimal exploitation of this nature, giant projects have come together in NEOM, such as the Trogena project, which will become a distinguished name in eco-tourism"

NEOM will be built around the concept of The Line, a mega project located in NEOM smart city, launched in January 2021. The Line will be a city with a million residents and a length of 170 km that preserves 95% of nature with zero cars, zero streets, and zero carbon emissions [260]. The Line concept has attracted much debate and interest due to its radically different approach towards urban living. It was detected as a separate cluster by our BERT model, but we merged it into one parameter called NEOM. The following are tweets related to this parameter:

"ذا لاين مدينة المستقبل في نيوم، المدينة تستهدف تحقيق مثالية العيش ومعالجة التحديات الملحة التي تواجه البشرية."

"The Line, the city of the future in NEOM, the city aims to achieve the ideal of living and address the urgent challenges facing humanity"

"ذا لاين احدى مشاريع المستقبل واللي يخدم البيئة ويرفع من جودة الحياة ويرفع من مستوى السياحة والاقتصاد المحلي"

"The Line is one of the future projects that serves the environment, raises the quality of life and raises the level of tourism and the local economy"

Significant Twitter activity was detected around the topic of the Neom International Airport. Our BERT model detected it as a separate cluster. We merged it with the NEOM for knowledge structure and simplicity. Below is an example tweet on the subject.

"انطلقت أولى الرحلات الدولية من مطار نيوم إلى دبي ، خطوة مميزة لتحقيق رؤية المملكة ...شكرا للهيئة العامة للطيران المدني وكافة الجهات"

"The first international flight departed from NEOM Airport to Dubai, an important step to achieve the Kingdom's vision... thanks to the General Authority of Civil Aviation and all"

5.3.4. AlUla City

The fourth parameter is AlUla City. The Saudi city of AlUla has become one of the new tourist destinations due to its beautiful natural components, diverse history, and antiquities dating back thousands of years. It includes the keywords AlUla, Umluj, Summit, Mada'in, Jeddah, Decisions, Higher, Gulf, Ministers, and Tabuk. The following tweets were posted:

"العلا أعظم مدينه تاريخيةفي المملكة العربية السعودية، وتوجد بها مدائن صالح و هي وجهة سياحية ر ائعةتتمتع بالعديدمن الثقافات والمواقعالأثرية"

"AlUla is the greatest historical city in the Kingdom of Saudi Arabia, and there is Mada'in Saleh, which is a wonderful tourist destination with many cultures and archaeological sites"

"مدائن صالح في #العلا أول موقع سعودي مُدرج على لائحة اليونسكو للتراث العالمي"

"Mada'in_Saleh in AlUla is the first Saudi site to be inscribed on the UNESCO World Heritage List".

"محافظة العلا هي تحفه فنيه رائع توجد فيها (مدائن صالح) احد الاثار قديمة والجميلة"

"AlUla Governorate is a wonderful masterpiece in which there is (Madain Saleh), one of the ancient and beautiful monuments"

5.3.5. Abha City

The fifth parameter is Abha City (also called the Abha Tourist City by Saudi people). The keywords of this parameter include Abha, Capital, Region, Achievement, Asir, and Tourism. The following are examples of tweets related to this topic:

"ابها البهية حباها الله بطبيعة خلابة وصيف ممتع تكتسي فيه سماءها بغيوم متراكمه تشاهدها وانت بأعالي منتزهاتها..."

"Abha Al-Bahiya. God has blessed it with a beautiful nature and an enjoyable summer in which its sky is covered with accumulated clouds. You watch it while you are in parks above the clouds"

"مدينة أبها هي عاصمة السياحة العربية لعام 2017م، و الذي يعتبر انجاز للمدينة ولمنطقة عسير بشكل عام"

"Abha is the capital of Arab tourism for the year 2017. Which is considered an achievement for the city and the Asir region in general"

"السياحة السعودية: جهود ر ائعة في صيف أبها لهذا العام من قبل هيئة تطوير عسير. نشاطات وفعاليات ومهرجانات متعدد ة استمتع زوار أبها بأجواء رائعة"

"Saudi Tourism: Great efforts in Abha summer this year by the Asir Development Authority Multiple activities, events, and festivals. Visitors to Abha enjoyed a wonderful atmosphere"

5.3.6. Jeddah Historical

The sixth parameter is Jeddah Historical, also called AlBalad, which means town. It is a historical district of Jeddah city in Saudi Arabia. This parameter is characterized by the keywords such as Historical, Tour, Jeddah, Host, Geographic, Its Position, Importance, Assistance, Helper, Geologic, and Land. The following are some tweets related to this parameter:

"زرت اليوم واستمعت برحلة في تاريخ جده القديمةو مواقعها الاثرية ، وجدت عدداً من السياح وعملت مترجماً لهم لما كتب على اللوحات"

"I visited today and listened to a journey in the history of ancient Jeddah and its archaeological sites, and I found a number of tourists and worked as a translator for them of what was written on the paintings"

"جده التاريخيه تعتبر من احد اهم الأماكن التاريخيه في دولتنا واهتمام ولي عهد في تطويرها يعكس حجمه"

"Historic Jeddah is considered one of the most important historical places in our country, and the interest of the Crown Prince in developing it reflects its importance"

5.3.7. Natural Places

The seventh parameter is Natural Places. The keywords of this parameter are Caucasus, Mountains, European, Nature, Paradise, Traveler, Touches, Mix, Dream, Wonderful, and Enjoy. Examples of the tweets regarding this parameter are below:

"تمتع بجمال الطبيعة في جبال القوقاز الخلابة بعد الغداء في مطاعم الساحات الاوروبية والعالمية . . . "

"Enjoy the beauty in spectacular Caucasus mountains after lunch in the restaurants of European and international squares . . . "

"زور عسير واستمتع باستكشاف الطبيعة الساحرةفي جبال السودةحتى قرية رجال ألمع التاريخية"

"visit Asir and enjoy exploring the enchanting nature in the Souda Mountains to the historic village of Rijal Alma"

"تروجينا . . . وجهة سياحية عالمية لعشاق الطبيعة والمغامرة والترفيه في أعالي جبال نيوم"

"Trogena . . . a global tourist destination for lovers of nature, adventure and entertainment in the high mountains of NEOM"

5.3.8. Seasonal Festivals

The eighth parameter is Seasonal Festivals. It is created by merging three clusters, 11, 16, and 17. The parameter captures various discussions about tourism related to seasonal entertainment festivals and events in Saudi Arabia. The keywords for the parameter include Tickets, Booking, Winterland, Season, Jeddah, Riyadh, Boulevard, Events, Jungle, Party, Music, Horrible, Fun, Show, and For Sale. For instance, the tweet outlined below was found in our dataset related to this parameter. It was posted by the booking company. This and similar tweets describe the Riyadh Season events and activities such as booking tickets and searching for flights to travel to Riyadh:

"احجز تذكرتك كافة انواع وفئات فعاليات موسم الرياض تذاكر حفلات ورحلات مناسبات والعاب ومنتزهات في البوليفارد . . . "

"Book your ticket for all types and categories of Riyadh Season events, tickets for parties, trips, events, games and parks on the Boulevard . . . "

5.3.9. Temporal Progression (Tourist Attractions)

Figure 22 plots the temporal progression of the macro-parameter Tourist Attractions. We generated the plots using the topic-over-time method in the BERTopic library. The graph's horizontal line represents the timeline of the discovered parameters during the data period of March 2021 to October 2022, while the vertical line represents the number of tweets, which is referred to as the intensity. The tweets related to popular tourist national and international destinations and some tourism-related development projects in Saudi Arabia. The National Destination parameter was the most discussed topic in 2022.

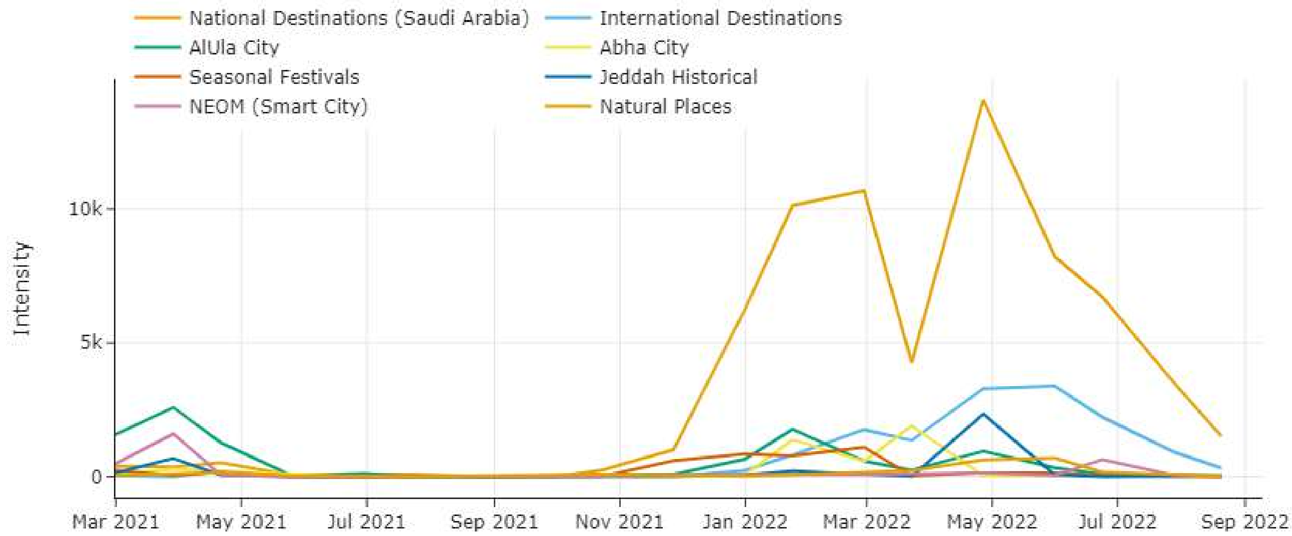


Figure 22. Temporal progression (macro-parameter: Tourist Attractions) (data source: Twitter).

5.4. Tourism Services

We discuss now the second macro-parameter, Tourism Services. It consists of five parameters. These are Programs, Tours, and Packages; Offers, Discounts, and Gifts; Tourist Guides; Restaurants and Hotels; and Medical Insurance and Internet Services.

5.4.1. Programs, Tours, and Packages

The first parameter in this macro-parameter (ninth overall) is Programs, Tours, and Packages. This parameter is created by merging three clusters, numbers 5, 15, and 27. Most tourists, when planning their trips, look for reasonable and suitable tourism programs. Usually, tourism and travel companies offer a variety of tourism programs. This parameter covers these issues and is represented by the keywords Tourist, Programs, Most Suitable, Hotel Reservation, Rental, Distinguished, Tours, Reservations, Apartments, Cars, Offers, and Families. The following tweets provide examples of the posts:

"[شركةسياحية] يقدم لك أفضل البرامج السياحية في منصة وحدة من قبل شركات تنظيم الرحلات"

"[Travel Company] offers you the best tourism programs in a single platform by tour operators".

"عيش المغامرة الان في جورجيا مع برامج سياحية مختلفة من [شركة سياحية] للسياحة والسفر"

"Live the adventure now in Georgia with different tourism programs from [Travel Company] for travel and tourism"

"وزارة الثقافة بالتعاون مع وزارة السياحة ينظمون لزوار الحرمين برامج سياحيه ثقافيه مصممه للمعتمرين زيارات، لاماكن دينيه رحلات لطريق لهجرة. زيارات لمواقع الغزوات"

"The Ministry of Culture, in cooperation with the Ministry of Tourism, organizes cultural tourism programs for visitors to . . . "

"السياحه في ... برامجنا تشمل شقق فاخره للعوائل والشباب استقبال وتوديع المطار فنادق ومنتجعات سيارة وسائق خاص رحلات يومي"

"Tourism in . . . Our programs include luxury apartments for families and youth Airport reception and farewell Hotels and resorts Car and private driver Daily trips"

The keywords related to Packages (Cluster 15) detected by our model include Camping, Accompany, Delivery, Reservation, Driver, Apartments, Guide, Airport, Prices, Tourists, Car, Flights, and Hotels. The following tweets provide examples of the posts:

"السياحه في ... برامجنا تشمل شقق فاخره للعوائل والشباب استقبال وتوديع المطار فنادق ومنتجعات سيارة وسائق خاص رحلات يومي"

"Tourism in . . . Our programs include luxury apartments for families and youth, Airport reception and farewell, Hotels and resorts, Car and private driver, daily trips"

Some examples of tweets related to Tours or trips are provided below.

"رحلات سياحية لمزارع إنتاج الورد في عروس المصائف الطائف في صيف_السعودية"

"Tourist trips to the rose production farms in the bride of summer resorts Taif City"

"هنوفر لك حجز فندقي أو شقق فندقية و سيارة خاصة و رحلات سياحية"

"We will provide you with a hotel reservation or hotel apartments, a private car and tourist trips"

"قريباً رحلات كروز السعودية تنطلق مجدداً من جدة إلى 3 وجهات في البحر الأحمر السعودية ومصر والأردن"

"Coming soon Saudi Cruise trips, it departs from Jeddah to three destinations in the Red Sea, Saudi Arabia Egypt Jordan"

5.4.2. Offers, Discounts, and Gifts

The next parameter is Offers, Discounts, and Gifts. This parameter is created by merging three clusters, numbers 29, 2, and 21. The keywords related to this parameter (in all clusters) are Offers, Travel, Book, Enjoy, Save, Prices, Qatar, Resort, Beauty, Tourism, Services, UAE, Azerbaijan, Special, Discounts, Companies, Entrepreneurship, Gifts, Big, Inquiries, Winter, Best, Electronic, Lovers, and Location. The following tweets provide examples of posts related to this parameter:

"الان أفضل عروض السفر و السياحة إلى جزيرة ... الجميلة سارع بالحجز الان على افضل عروض السياحه وسفر..."

"Now the best travel and tourism offers to the beautiful island . . . , hurry up to book now for the best tourism and . . . travel"

"سياحة في جورجيا خصومات خاصة للأعداد الكبيرة وخصومات خاصة للشركات وهدايا للعرسان الجداد لأحلي شهر عسل"

"Tourism in Georgia Special discounts for large numbers, special discounts for companies and gifts for newlyweds for the sweetest honeymoon"

"خصومات وعروض ومبادرات خاصة بالمواطنين لتشجيع السياحة الداخلية"

"Discounts, offers and initiatives for citizens to encourage domestic tourism"

"استمتع باكتشاف جمال وجهات صيفالسعودية بخصومات حصرية تصل حتى 50%"

"Enjoy discovering the beauty of Saudi summer destinations with exclusive discounts of up to 50%"

"عروض_سياحية خصومات مميزة علي اسعار تذاكر الطيران علي احصل 30%من سعر تذاكر سفرك..."

"Tourist Offers Special discounts on airline ticket prices. Get a 30% discount on the price of your travel tickets . . . "

5.4.3. Tourist Guides

The third parameter in this macro-parameter (eleventh overall) is Tourist Guides. It includes the following keywords: Inquiries, Guide, Your Trip, Offers, Reservations, Emirates, Bahrain, Oman, Tbilisi, Qatar, Tourism, Hotels, Areas, and Car. The following tweets provide examples of the posts:

"نسعد بخدمتكم في مجال الارشاد السياحي كمرشد سياحي في منطقة تبوك او نيوم اومحافظة العلا"

"We are pleased to serve you . . . as a tourist guide in the Tabuk, Neom, or AlUla regions".

"المرشد السياحي واجهه مشرفه لبلاده فقد تم تأهيله في مهارات الارشاد السياحي والاسعافات الاولية ... بإشراف مباشر من وزارة السياحة ...فهو يقدم المعلومة الصحيحة عن المكان والزمان وخير معين في السفر"

"The tour guide is qualified with the skills of tourist guidance and first aid ... managed by the Ministry of Tourism ... [tour guide] provides the correct information about the place and time and other information needed in travel"

"سائق ومرشد سياحي في اندونيسيايتكلم الانجليزية والعربية"

"Driver and tour guide in Indonesia who speaks English and Arabic"

"رحلة سياحية لزيارة مرتفعات الشفا، تتضمن نشاط الهايكنج بمرافقة مرشد سياحي متخصص"

"Tourist trip from Shefa Heights, hiking activity, accompanied by a specialized tour guide"

5.4.4. Restaurants and Hotels

Restaurants and Hotels constitute the fourth parameter (overall twelfth) for the macro parameter Tourism Services. This parameter is created by merging three clusters, numbers 23, 24, and 7. It highlights the issues and importance of restaurants and accommodation services in tourism. The keywords include Places, Hotels, Restaurants, Resorts, Tourist, Nearest, Finest, Offer, Country, Vacation, Sites, Cities, Visitor, Family, Winter, Bahrain, and UAE. Following are example tweets:

"في نادي_جدة_لليخوت استمتعوا بتجربة ممتعة مع المطاعم المطلة على البحر"

"In Jeddah yacht club, have an enjoyable experience with restaurants overlooking the sea"

"أفضل اطلالات و أفخم الفنادق في جورجيا بأسعار خاصة"

"The best views and the most luxurious hotels in Georgia at special prices"

"لتعزيز قطاع السياحة وتطوير وجهات سياحية نوعية، وقع صندوق التنمية السياحي مع مجموعة فنادق ومنتجعات... مذكرة تفاهم لشراكة استراتيجية تطويريه لجذب الاستثمار السياحي بالمملكة"

"To promote the tourism sector and develop quality tourist destinations, Tourism Development Fund signed up with a group of hotels and resorts a memorandum of understanding for a strategic development partnership to attract tourism investments in the Kingdom"

"تُقدّم العلا مجموعة متنوعة من أماكن الإقامة والمنتجعات الراقية"

"AlUla offers a variety of high-end accommodation and resorts"

"تمتع بأجواء الاقامة فى افخم المنتجعات والفنادق والأكواخ في جورجيا"

"Enjoy the atmosphere of accommodation in the most luxurious resorts, hotels, and cottages in Georgia"

5.4.5. Medical Insurance and Internet Services (MI2S)

The thirteenth parameter is Medical Insurance and Internet Services (MI2S). The global tourism industry has faced significant difficulties because of the COVID-19 pandemic. The requirements for traveling during and after the pandemic have been altered by policies and travel restrictions established in numerous countries. Travel insurance is important to protect from potential health hazards and financial losses. The keywords related to this parameter are Insurance, Get, Trip, SIM Cards, Internet, Driver, Taxi, Published, Tourist, Company, Best, Insurance, and Travel. The reason for the two topics—insurance and Internet services—is some companies offer the two services together.

"احجز رحلتك معنا وأحصل على تأمين السفر وشرائح انترنت مجاناوخصم خاص للحجزالمسبق..."

"Book your trip with us and get travel insurance, free internet SIM cards, and a special discount for pre-booking for reservations..."

"الخطوط السعودية تشترط تأمين طبي للأطفال"

"Saudi Airlines requires medical insurance for children"

The following tweet was posted by Saudi Tourism Authority indicate the importance of obtaining medical insurance to cover COVID-19 risk when arriving at Saudi airports.

"يتعين على كافة الزوار الحصول على تأمين طبي ضد مخاطر كورونا؛ والتي يمكن الحصول عليها من جميع المطارات الدولية في المملكة عند الوصول"

"All visitors are required to have medical insurance against corona risks, which can be obtained from all international airports in the Kingdom upon arrival"

5.4.6. Temporal Progression (Tourism Services)

The temporal progression of the macro-parameter Tourism Services, which includes five parameters, is shown in Figure 23. A mix of behaviour can be observed for all parameters. Overall, the Offers, Discounts, and Gifts parameter shows the highest intensity among all, and this could be due to businesses posting tweets about their offers.

Figure 23. Temporal progression (macro-parameter: Tourism Services) (data source: Twitter).

6. Discussion

The aim of this paper is to gain a comprehensive understanding of tourism and develop a theory and approach for smarter, sustainable tourism through the use of cutting-edge technologies. The study modelled the tourism industry using scientific research papers and tweets. It used a data-driven approach with deep learning and big data to extract parameters from both academic literature and public opinions on Twitter to give a comprehensive view of the industry from two different perspectives. A software tool for smart tourism was developed with four modules using BERT embeddings, UMAP, HDBSCAN, and TF-IDF, grouping discovered parameters into macro-parameters, validated internally and externally, and visualized with Seaborn, Plotly, and Matplotlib libraries. The paper also presented a comprehensive knowledge structure and literature review of the tourism sector, drawing on more than 250 research articles.

The academic-view dataset was constructed using 156,759 English research articles from the Scopus database covering 2000–2022. It was used to uncover key elements of academia-oriented tourism. Thirty-three parameters related to tourism were identified and grouped into four categories: Tourism Types, Tourism Planning, Tourism Challenges, and Media and Technologies. The "Tourism Types" macro-parameter discovered that the tourism industry encompasses a variety of activities and experiences such as urban, rural, beach and marine, national parks, wildlife, adventure, diving, sports, space, culture and heritage, music and art, and religious tourism. Activities range from city sightseeing, rural life experience, swimming, snorkeling, safari, adventure, visiting museums and historical sites, attending concerts, visiting art galleries, and religious tourism. The "Tourism Planning" macro-parameter captured details of planning for tourism that involves investing in education and training, utilizing air transport, and considering the hotel and food service industries to drive sustainable economic growth in communities. Forecasting methods are used to predict future demand. It also helps to develop underdeveloped communities by providing jobs and income opportunities.

The "Tourism Challenges" parameter captured several challenges that the tourism industry is facing such as climate change, poor air and water quality, natural disasters, and pandemics. Climate change is affecting the availability of tourist destinations and activities. Poor air and water quality can make tourist destinations less attractive and have negative effects on visitors' health. Natural disasters and pandemics can damage tourism infrastructure and disrupt travel plans. To overcome these challenges, research in sustainable tourism is underway such as reducing carbon emissions, improving waste management, and promoting sustainable transportation options, which will minimize negative impacts on the environment and local communities while promoting the long-term viability of the tourism industry. The "Media & Technology" parameter captured the major impact that media and technology are having on the tourism industry. Film and TV have been used traditionally to showcase destinations. Online and social media platforms have now become essential for promotion. Mobile apps have made it easier for travelers to plan and book trips, and virtual reality technology allows people to experience a destination before visiting or enhance their experiences during visits.

The Twitter dataset for this study was collected for 18 months from March 2021 to August 2022 and consisted of 485,813 tweets related to the public perception of tourism in Saudi Arabia. The dataset was limited to the region to focus on local tourism issues and was modelled to reveal 13 parameters, grouped into two broader categories: Tourist Attractions and Tourism Services.

The "Tourist Attractions" parameter captured a diverse range of tourist attractions that Saudi Arabia offers for both domestic and international travelers, including popular National destinations such as the NEOM Smart City, AlUla City, Abha City, and Jeddah, and many natural places such as deserts, mountains, and beaches. Additionally, Saudi Arabia hosts many seasonal festivals that allow tourists to experience the country's culture and traditions. With a variety of options catering to different interests, Saudi Arabia is an attractive destination for tourists. The parameter also captured many international destinations that are popular for Saudi tourists including, Qabil, Qubam and Baku (Azerbaijan), Batumi, Tbilisi, and Borjomi (Georgia), and Dubai, Emirates, Qatar, Bahrain, Bosnia, and Trabzon (Tukiye). The "Tourism Services" captured a wide range of tourism services Saudi Arabia offers for both domestic and international travelers, including tour packages, discounts and gifts, and tourist guides. They also offer a variety of hotels and restaurants and options for medical insurance and internet services.

The two perspectives, academic versus public, or international versus national, are not isolated and have some impact on each other, but they still have distinct and significant variations. In Figure 1, in comparing these two perspectives, we presented a multi-perspective taxonomy of the tourism sector, combining academic, public, international, and national/cultural (Saudi Arabia) perspectives. It offered a holistic understanding of the industry, including 15 types of tourism, various planning dimensions, major challenges, and the impact of media and technology (academic and international view). The national/public perspective in Saudi Arabia focuses on tourist attractions and services such as medical insurance, including recent developments such as the NEOM smart city, AlUla city, and seasonal festivals.

Figure 24 presents a data-driven framework for smarter tourism, aimed at improving tourist experiences and promoting sustainable practices. The framework is based on data sources such as social media, academic articles, government, and industry, among others. The objectives include improving the quality of services, cultural sustainability, economic sustainability, environmental sustainability, and tourist experiences. Enablers for achieving these objectives include tourism satisfaction, efficiency, experiences, and emerging technologies. The framework also lists challenges in tourism, such as climate change, air pollution, water quality, security, lack of awareness, and food safety. These challenges should be addressed by the framework to achieve the desired goals.

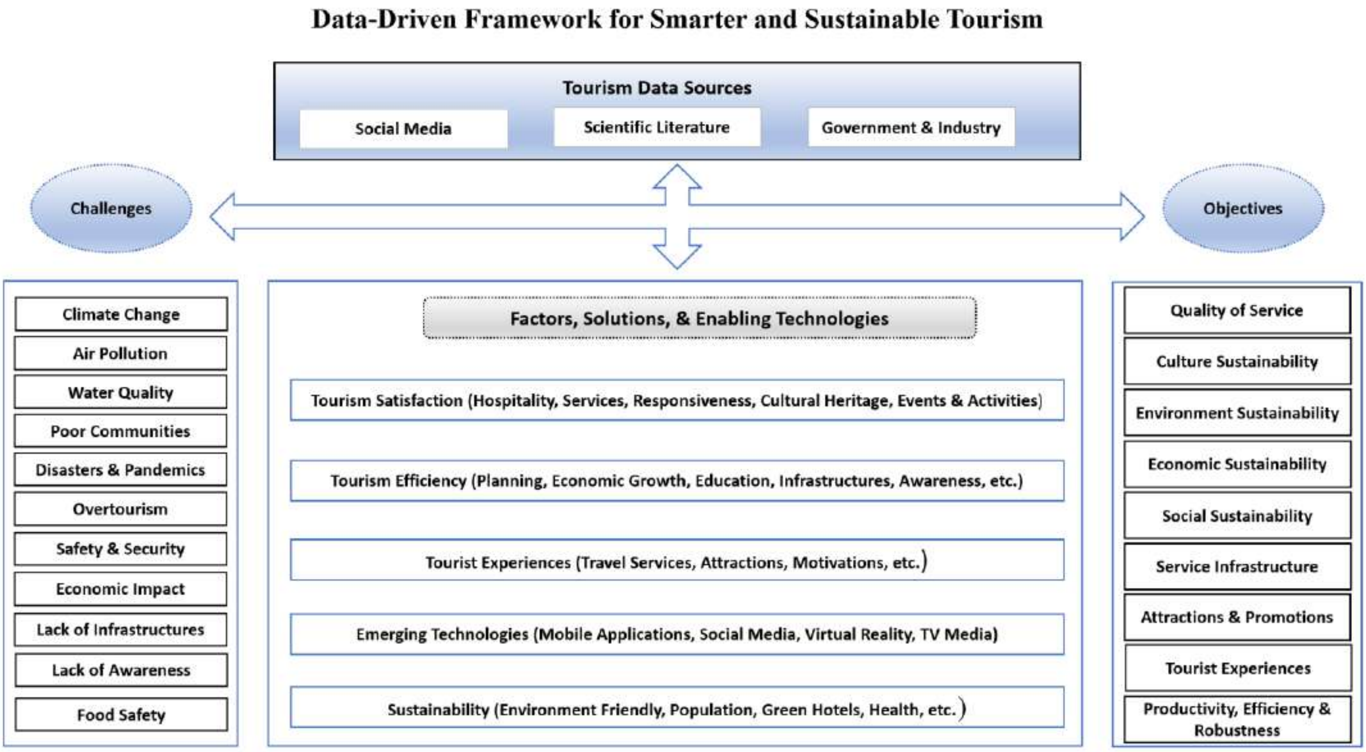


Figure 24. Data-driven framework for smarter sustainable tourism.

6.1. Novelty and Utilization

The tourism industry is constantly evolving and facing new challenges such as emerging technologies, global conflicts, energy and monetary crises, pandemics, and disasters. A literature review (see Section 2) and an extensive analysis of current research (see Section 4) show that the field is fragmented and narrowly focused. To effectively navigate and improve this dynamic sector, a holistic approach is needed to study tourism. A holistic view of the tourism industry considers all aspects of the industry, including economic, social, environmental, and cultural factors. By taking this comprehensive approach, decision-makers can gain a deeper understanding of the interconnectedness of different aspects of tourism and how they impact one another. This knowledge can be used to drive sustainable tourism practices and destination development and enhance visitor experiences. Additionally, a holistic approach can help identify potential challenges and opportunities in advance, allowing for proactive rather than reactive solutions. This proactive approach is crucial in today's fast-paced environment as it helps to stay ahead of industry trends and mitigate potential negative impacts. Furthermore, a holistic approach to tourism can have a positive impact on both tourists and local communities, promoting sustainable tourism practices that benefit the environment and local communities and creating better visitor experiences that lead to repeat visits and positive word of mouth.

Moreover, the relationship and effects of tourism on local cultures in the emerging digital world have also attracted limited attention, particularly in Saudi Arabia. In these contexts, research on big data analytics of social and digital media, particularly in the Arabic language, is limited. Saudi Arabia is a rapidly developing country with a unique culture and a diverse range of tourist attractions. Conducting tourism research in this country can provide insight into the latest trends and developments in the industry and help to identify potential opportunities for tourism growth. It can also allow researchers to explore the country's attractions and learn more about its culture and history.

This paper presents an approach for leveraging machine and deep learning to gather holistic, multi-perspective (e.g., local, cultural, national, and international), and objective information on any subject including tourism. By providing powerful tools and resources to analyze various datasets, it makes it possible to uncover crucial information on matters of public, academic, industrial, and government interest. The research and insights discovered in this paper make a significant contribution to our understanding of the tourism industry and have the potential to shape public perception, guide future research, and inform decision-making by the public, the government, and other stakeholders.

The proposed approach not only enhances the theory and practice of AI-based methods for information discovery but also extends the use of the scientific literature, Twitter, and other media and data sources for information and parameter discovery to enable autonomous capabilities for holistic and dynamic optimizations in everyday applications, systems, and platforms. Furthermore, it promotes novel approaches to research in the tourism sector using the information discovery approach, ultimately giving rise to the development of smart and sustainable societies, economies, and the planet, which is a paramount concern for today's world.

As technology advances, more and more systems such as self-driving cars, web services, drones, and robots in manufacturing and farming are becoming autonomous. This trend is likely to continue, and we will see this kind of autonomous functionality being incorporated into an even wider variety of systems, including those used in industry, city, and country management. Even when a system is not fully autonomous, understanding its parameters is still important, as they form the basis for decision-making and problem-solving in the design and operation of the system.

7. Conclusions

The fragility of the tourism industry, which prior to the COVID-19 pandemic contributed 10.3% to the global GDP and employed 333 million people, is being exposed by global natural and manmade events and requires collaboration and comprehensive understanding for responsible and innovative growth towards sustainable and smart tourism. The aim of this study is to gain a comprehensive understanding of tourism, drive future research through cutting-edge technologies (artificial intelligence, big data, and others), and ultimately develop a theory and approach for smarter tourism that supports sustainable future societies. This paper presented a machine learning approach to extract parameters from the academic literature and public opinions on Twitter to provide a holistic view of the tourism industry and promote sustainable and smart tourism through improved AI-based information discovery. The approach modelled 156,759 research articles and 485,813 tweets to identify 33 distinct parameters in 4 categories for the academic perspective and 13 parameters for public perception. The paper presents a comprehensive knowledge structure and literature review of the tourism sector, drawing on more than 250 research articles.

7.1. Theoretical and Practical Implications

The work presented in this paper has significant theoretical and practical implications for the tourism industry and beyond and is of critical importance in the current rapidly evolving technological landscape (see also Section 6.1). By highlighting the need for a comprehensive, holistic approach to studying the tourism industry, the paper provides a framework that can be used to navigate the complex challenges facing the industry, such as emerging technologies, global conflicts, pandemics, and disasters.

The practical implications of this work are numerous and far-reaching, with the development of powerful tools and resources for analyzing diverse datasets representing a major step forward in the effective management and optimization of complex systems. These tools and resources are critical in today's fast-paced environment, where proactive rather than reactive solutions are needed to stay ahead of industry trends and mitigate potential negative impacts.

Additionally, the proposed approach has the potential to extend the use of scientific literature, Twitter, and other media and data sources for information and parameter discovery to enable autonomous capabilities for holistic and dynamic optimizations in everyday applications, systems, and platforms. By leveraging machine and deep learning to gather objective and multi-perspective information, decision-makers can make more informed choices, leading to the development of smart and sustainable societies, economies, and the planet.

Overall, this work represents a significant contribution to our understanding of the tourism industry and has the potential to shape public perception, guide future research, and inform decision-making by the public, government, and other stakeholders. It is essential reading for anyone interested in the future of the tourism industry and the development of sustainable, technology-driven solutions to the challenges facing our planet.

7.2. Limitations

While the work presented in this paper is undoubtedly important and valuable, it is important to acknowledge its limitations. One of the primary limitations is that the study focuses on the tourism industry in Saudi Arabia, and while this is a rapidly developing country with unique cultural and tourist attractions, the findings may not be applicable to other countries or regions.

Additionally, the proposed approach for leveraging machine and deep learning to gather objective and multi-perspective information has its own limitations. For example, the accuracy and reliability of the data sources used to train the machine learning algorithms can affect the quality of the insights generated. Similarly, the complexity of the algorithms used can make it difficult to interpret the results and identify potential biases.

Another limitation is that the approach proposed in this paper is heavily reliant on technological infrastructure, which may not be available or accessible in all regions. This could limit the applicability of the approach in certain contexts and may prevent some decision-makers from accessing the insights and resources that it provides.

7.3. Future Work

The paper belongs to our extensive research on utilizing Information and Communication Technology (ICT) to tackle issues in smart cities and societies. Our work encompasses deep journalism [51,261], labor economics [262], transportation [51], smart families and homes [52], healthcare [62,263], education during COVID-19 [64], and event detection [67]. In the future, we aim to enhance the methodology in this paper through advanced deep learning techniques and apply them to enhancing tourism and other societal, economic, environmental, and cultural issues. This research utilized Scopus database and Twitter data to uncover parameters. In the future, we plan to integrate other scientific databases, social and digital media, and additional data sources to expand the scope of our findings and provide a more comprehensive understanding. Finally, we note that the development of sustainable tourism must necessarily be based on intangible factors of territorial development, such as intellectual capital in particular, such as noted in [264,265]. Future work will also look into this direction.

Author Contributions: Conceptualization, R.A. and R.M.; methodology, R.A. and R.M.; software, R.A.; validation, R.A. and R.M.; formal analysis, R.A., R.M. and A.A.; investigation, R.A., R.M. and A.A.; resources, R.M. and A.A.; data curation, R.A.; writing—original draft preparation, R.A. and R.M.; writing—review and editing, R.M. and A.A.; visualization, R.A.; supervision, R.M. and A.A.; project administration, R.M.; funding acquisition, R.M. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge with thanks the technical and financial support from the Deanship of Scientific Research (DSR) at the King Abdulaziz University (KAU), Jeddah, Saudi Arabia, under Grant No. RG-11-611-40.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this paper can be made available subject to the data providers' terms and conditions.

Acknowledgments: The work carried out in this paper is supported by the HPC Center at the King Abdulaziz University. The training and software development work reported in this paper was carried out on the Aziz supercomputer.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

BERT	Bidirectional Encoder Representations from Transformer
HDBSCAN	Hierarchical Density-Based Spatial Clustering of Applications with Noise
UMPA	Uniform Manifold Approximation and Projection
TF–IDF	Term Frequency-Inverse Document Frequency
ICTs	Information and Communication Technologies
UNWTO	World Tourism Organization
AI	Artificial intelligence
NLTK	Natural Language Toolkit
NLP	Natural language processing
EGUC	Economic Growth for Underdeveloped Communities
WQPM	Water Quality & Pollution Management
MI2S	Medical Insurance & Internet Services

References

1. Del Vecchio, P.; Mele, G.; Ndou, V.; Secundo, G. Creating Value from Social Big Data: Implications for Smart Tourism Destinations. *Inf. Process. Manag.* **2018**, *54*, 847–860. [CrossRef]
2. Kontogianni, A.; Alepis, E. Smart Tourism: State of the Art and Literature Review for the Last Six Years. *Array* **2020**, *6*, 100020. [CrossRef]
3. Mehmood, R.; See, S.; Katib, I.; Chlamtac, I. *Smart Infrastructure and Applications: Foundations for Smarter Cities and Societies*; Springer International Publishing: Cham, Switzerland; Springer Nature: Cham, Switzerland, 2020; ISBN 9783030137045.
4. Mehmood, R.; Sheikh, A.; Catlett, C.; Chlamtac, I. Editorial: Smart Societies, Infrastructure, Systems, Technologies, and Applications. *Mob. Netw. Appl.* **2022**, *1*, 1–5. [CrossRef]
5. Zhuang, X.; Liu, H. Some Suggestions for Community-Based Ecotourism Management. In Proceedings of the 2010 International Conference on Management and Service Science, Wuhan, China, 24–26 August 2010. [CrossRef]
6. Li, J.; Lee, T.J.; Chen, N.; Park, K.-S. Pro-Environmental Behaviour of the Residents in Sensitive Tourism Destinations. *J. Vacat. Mark.* **2022**. [CrossRef]
7. United Nations Department of Economic and Social Affairs Sustainable Tourism. Available online: https://sdgs.un.org/topics/sustainable-tourism (accessed on 26 November 2022).
8. Nekmahmud, M.; Ramkissoon, H.; Fekete-Farkas, M. Green Purchase and Sustainable Consumption: A Comparative Study between European and Non-European Tourists. *Tour. Manag. Perspect.* **2022**, *43*, 100980. [CrossRef]
9. Schubert, S.F.; Schamel, G. Sustainable Tourism Development: A Dynamic Model Incorporating Resident Spillovers. *Tour. Econ.* **2021**, *27*, 1561–1587. [CrossRef]
10. Wang, S.; Wang, J.; Li, J.; Yang, F. Do Motivations Contribute to Local Residents' Engagement in pro-Environmental Behaviors? Resident-Destination Relationship and pro-Environmental Climate Perspective. *J. Sustain. Tour.* **2020**, *28*, 834–852. [CrossRef]
11. Saxon, S.; Sodprasert, J.; Sucharitakul, V. Reimagining Travel: Thailand Tourism after the COVID-19 Pandemic. Available online: https://www.mckinsey.com/industries/travel-logistics-and-infrastructure/our-insights/reimagining-travel-thailand-tourism-after-the-covid-19-pandemic (accessed on 1 February 2023).
12. World Travel & Tourism Council (WTTC) Travel & Tourism Economic Impact. Available online: https://wttc.org/research/economic-impact (accessed on 20 December 2022).
13. Global Tourism Industry—Statistics & Facts | Statista. Available online: https://www.statista.com/topics/962/global-tourism/#topicOverview (accessed on 6 January 2023).

14. Ali, A. Travel and Tourism: Growth Potentials and Contribution to the GDP of Saudi Arabia. *Probl. Perspect. Manag.* **2018**, *16*, 417–427. [CrossRef]
15. Arabian Business. Available online: https://arabianbusiness.com (accessed on 30 April 2019).
16. Mocák, P.; Matlovičová, K.; Matlovič, R.; Pénzes, J.; Pachura, P.; Mishra, P.K.; Kostilníková, K.; Demková, M. 15-Minute City Concept as a Sustainable Urban Development Alternative: A Brief Outline of Conceptual Frameworks and Slovak Cities as a Case. *Folia Geogr.* **2022**, *64*, 69–89.
17. Tripathy, A.K.; Tripathy, P.K.; Ray, N.K.; Mohanty, S.P. ITour: The Future of Smart Tourism. *IEEE Consum. Electron. Mag.* **2018**, *7*, 32–37. [CrossRef]
18. Afzaal, M.; Usman, M.; Fong, A. Predictive Aspect-Based Sentiment Classification of Online Tourist Reviews. *J. Inf. Sci.* **2019**, *45*, 341–363. [CrossRef]
19. Loureiro, S.M.C.; Guerreiro, J.; Ali, F. 20 Years of Research on Virtual Reality and Augmented Reality in Tourism Context: A Text-Mining Approach. *Tour. Manag.* **2020**, *77*, 104028. [CrossRef]
20. Karas, Y.K.; Atay, L.; Sevinc, H.K.; Duru, A. Investigation of GIS Based Mobile Route Planning-Navigation Applications for Tourism Activities. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences—ISPRS Archives, Aksaray, Turkey, 20 December 2021; Volume 46, pp. 321–326.
21. Li, J.; Xu, L.; Tang, L.; Wang, S.; Li, L. Big Data in Tourism Research: A Literature Review. *Tour. Manag.* **2018**, *68*, 301–323. [CrossRef]
22. Rashideh, W. Blockchain Technology Framework: Current and Future Perspectives for the Tourism Industry. *Tour. Manag.* **2020**, *80*, 104125. [CrossRef]
23. Li, H.; Hu, M.; Li, G. Forecasting Tourism Demand with Multisource Big Data. *Ann. Tour. Res.* **2020**, *83*, 102912. [CrossRef]
24. Fronzetti Colladon, A.; Guardabascio, B.; Innarella, R. Using Social Network and Semantic Analysis to Analyze Online Travel Forums and Forecast Tourism Demand. *Decis. Support Syst.* **2019**, *123*, 113075. [CrossRef]
25. Serrano, L.; Ariza-Montes, A.; Nader, M.; Sianes, A.; Law, R. Exploring Preferences and Sustainable Attitudes of Airbnb Green Users in the Review Comments and Ratings: A Text Mining Approach. *J. Sustain. Tour.* **2021**, *29*, 1134–1152. [CrossRef]
26. Chen, W.; Xu, Z.; Zheng, X.; Yu, Q.; Luo, Y. Research on Sentiment Classification of Online Travel Review Text. *Appl. Sci.* **2020**, *10*, 5275. [CrossRef]
27. Miah, S.J.; Vu, H.Q.; Gammack, J.; McGrath, M. A Big Data Analytics Method for Tourist Behaviour Analysis. *Inf. Manag.* **2017**, *54*, 771–785. [CrossRef]
28. Wang, L.; Wang, X.K.; Peng, J.J.; Wang, J. qiang The Differences in Hotel Selection among Various Types of Travellers: A Comparative Analysis with a Useful Bounded Rationality Behavioural Decision Support Model. *Tour. Manag.* **2020**, *76*, 103961. [CrossRef]
29. Chao, C.L.; Lin, K.C.; Tseng, C.M.; Erdene, T.; Chen, Y.F. An App for Promoting Health and Local Tourism. In Proceedings of the 2014 International Symposium on Computer, Consumer and Control, Taichung, Taiwan, 10–12 June 2014; pp. 1175–1178. [CrossRef]
30. Bres, S.; Tellez, B. Localisation and Augmented Reality for Mobile Applications in Culture Heritage. 2006; pp. 1–5. Available online: https://www.isprs.org/proceedings/xxxviii/5-w1/pdf/bres_tellez.pdf (accessed on 1 February 2023).
31. Jiang, S.; Moyle, B.; Yung, R.; Tao, L.; Scott, N. Augmented Reality and the Enhancement of Memorable Tourism Experiences at Heritage Sites. *Curr. Issues Tour.* **2022**, *26*, 242–257. [CrossRef]
32. Slini, T.; Papakostas, K.T. 30 Years Air Temperature Data Analysis in Athens and Thessaloniki, Greece. *Green Energy Technol.* **2016**, *Part F2*, 21–33. [CrossRef]
33. Tianning, Y.; Ma, X. Cost-Effectiveness Analysis of Greenhouse Dehumidification and Integrated Pest Management Using Air's Water Holding Capacity-a Case Study of the Trella Greenhouse in Taizhou, China. *E3S Web Conf.* **2021**, *251*, 2063. [CrossRef]
34. Hernández-Travieso, J.G.; Ravelo-García, A.G.; Alonso-Hernández, J.B.; Travieso-González, C.M. Neural Networks Fusion for Temperature Forecasting. *Neural Comput. Appl.* **2018**, *32*, 15699–15710. [CrossRef]
35. Younes, M.B.; Boukerche, A. Safety Traffic Speed Recommendations for Critical Road Scenarios Using Vehicular Networks. In Proceedings of the 2018 IEEE International Conference on Communications (ICC), Kansas City, MO, USA, 20–24 May 2018. [CrossRef]
36. Koens, K.; Klijs, J. Overtourism—Identifying the Underlying Causes and Tensions in European Tourism Destinations. In *A Research Agenda for Urban Tourism*; Edward Elgar Publishing: Cheltenham, UK, 2022; pp. 243–256. [CrossRef]
37. Perkumien, D.; Pransk, R. Overtourism: Between the Right to Travel and Residents' Rights. *Sustainability* **2019**, *11*, 2138. [CrossRef]
38. Capocchi, A.; Vallone, C.; Pierotti, M.; Amaduzzi, A. Overtourism: A Literature Review to Assess Implications and Future Perspectives. *Sustainability* **2019**, *11*, 3303. [CrossRef]
39. Dodds, R.; Butler, R. The Phenomena of Overtourism: A Review. *Int. J. Tour. Cities* **2019**, *5*, 519–528. [CrossRef]
40. Koens, K.; Postma, A.; Papp, B. Is Overtourism Overused? Understanding the Impact of Tourism in a City Context. *Sustainability* **2018**, *10*, 4384. [CrossRef]
41. Dzhusibalieva, A.; Mauina, G.; Kasimbekova, M.; Serikbayeva, S.; Balginova, K.; Zhanibekova, G.; Dusembaeva, L.; Tasmambetova, A.; Bekzhanova, T.; Aktymbayeva, A.; et al. Prospects for a Green Economy in Tourism. *J. Internet Bank. Commer.* **2016**, *21*.

42. Matlovičová, K.; Mocák, P.; Kolesárová, J. Environment of Estates and Crime Prevention through Urban Environment Formation and Modification. *Geogr. Pannon.* **2016**, *20*, 168–180. [CrossRef]
43. Pandey, D.K.; Kumar, R. Russia-Ukraine War and the Global Tourism Sector: A 13-Day Tale. *Curr. Issues Tour.* **2022**, *26*, 692–700. [CrossRef]
44. Balli, F.; Billah, M.; Chowdhury, I. Impact of the Russia–Ukraine War on Hospitality Equity Markets. *Tour. Econ.* **2022**, 13548166221133493. [CrossRef]
45. Arfat, Y.; Usman, S.; Mehmood, R.; Katib, I. Big Data for Smart Infrastructure Design: Opportunities and Challenges. In *Smart Infrastructure and Applications Foundations for Smarter Cities and Societies*; Springer: Cham, Switzerland, 2020; pp. 491–518.
46. Alam, F.; Mehmood, R.; Katib, I.; Albogami, N.N.; Albeshri, A. Data Fusion and IoT for Smart Ubiquitous Environments: A Survey. *IEEE Access* **2017**, *5*, 9533–9554. [CrossRef]
47. Yigitcanlar, T.; Butler, L.; Windle, E.; Desouza, K.C.; Mehmood, R.; Corchado, J.M. Can Building "Artificially Intelligent Cities" Safeguard Humanity from Natural Disasters, Pandemics, and Other Catastrophes? An Urban Scholar's Perspective. *Sensors* **2020**, *20*, 2988. [CrossRef]
48. Baker, H.K.; Kumar, S.; Pattnaik, D. Twenty-Five Years of the Journal of Corporate Finance: A Scientometric Analysis. *J. Corp. Financ.* **2021**, *66*, 101572. [CrossRef]
49. Zhong, B.; Wu, H.; Li, H.; Sepasgozar, S.; Luo, H.; He, L. A Scientometric Analysis and Critical Review of Construction Related Ontology Research. *Autom. Constr.* **2019**, *101*, 17–31. [CrossRef]
50. Heilig, L.; Voß, S. A Scientometric Analysis of Public Transport Research. *J. Public Transp.* **2015**, *18*, 111–141. [CrossRef]
51. Ahmad, I.; Alqurashi, F.; Abozinadah, E.; Mehmood, R. Deep Journalism and DeepJournal V1.0: A Data-Driven Deep Learning Approach to Discover Parameters for Transportation. *Sustainability* **2022**, *14*, 5711. [CrossRef]
52. Alqahtani, E.; Janbi, N.; Sharaf, S.; Mehmood, R. Smart Homes and Families to Enable Sustainable Societies: A Data-Driven Approach for Multi-Perspective Parameter Discovery Using BERT Modelling. *Sustainability* **2022**, *14*, 13534. [CrossRef]
53. Alsaigh, R.; Mehmood, R.; Katib, I. AI Explainability and Governance in Smart Energy Systems: A Review. *arXiv* **2022**, arXiv:2211.00069. [CrossRef]
54. Mustak, M.; Salminen, J.; Plé, L.; Wirtz, J. Artificial Intelligence in Marketing: Topic Modeling, Scientometric Analysis, and Research Agenda. *J. Bus. Res.* **2021**, *124*, 389–404. [CrossRef]
55. Zach, F.J.; Krizaj, D.; Pretnar, A. Topic Modelling Tourism Literature on Innovation and Technology. *e-Rev. Tour. Res.* **2019**, *17*, 317–333.
56. Barrera-Barrera, R. Selecting the Appropriate Leading Journal in Hospitality and Tourism Research: A Guide Based on the Topic-Journal Fit and the JCR Impact Factor. *Scientometrics* **2022**, *127*, 1801–1823. [CrossRef]
57. Fang, Y.; Yin, J.; Wu, B. Climate Change and Tourism: A Scientometric Analysis Using CiteSpace. *J. Sustain. Tour.* **2017**, *26*, 108–126. [CrossRef]
58. Chen, J.; Zhou, W. The Exploration of Travel Motivation Research: A Scientometric Analysis Based on CiteSpace. *COLLNET J. Scientometr. Inf. Manag.* **2021**, *14*, 257–283. [CrossRef]
59. Ribeiro, M.I.B.; Fernandes, A.J.G.; Lopes, I.M. Smart Tourism: A Bibliometric Analysis of Scientific Publications from the Scopus and Web of Science Databases. *Smart Innov. Syst. Technol.* **2021**, *209*, 1–14. [CrossRef]
60. Baqeri, H.; Dehghan Dehnavi, H.; Nayebzadeh, S.; Rabbani, M.; Honari, M.T. Comprehensive Scientometrics Analysis of International Research in the Field of Mining Tourism Potential. *Int. J. Nonlinear Anal. Appl.* **2022**, 2008–6822. [CrossRef]
61. Alotaibi, S.; Mehmood, R.; Katib, I.; Rana, O.; Albeshri, A. Sehaa: A Big Data Analytics Tool for Healthcare Symptoms and Diseases Detection Using Twitter, Apache Spark, and Machine Learning. *Appl. Sci.* **2020**, *10*, 1398. [CrossRef]
62. Alahmari, N.; Alswedani, S.; Alzahrani, A.; Katib, I.; Albeshri, A.; Mehmood, R. Musawah: A Data-Driven AI Approach and Tool to Co-Create Healthcare Services with a Case Study on Cancer Disease in Saudi Arabia. *Sustainability* **2022**, *14*, 3313. [CrossRef]
63. Yigitcanlar, T.; Regona, M.; Kankanamge, N.; Mehmood, R.; D'Costa, J.; Lindsay, S.; Nelson, S.; Brhane, A. Detecting Natural Hazard-Related Disaster Impacts with Social Media Analytics: The Case of Australian States and Territories. *Sustainability* **2022**, *14*, 810. [CrossRef]
64. Alswedani, S.; Mehmood, R.; Katib, I. Sustainable Participatory Governance: Data-Driven Discovery of Parameters for Planning Online and In-Class Education in Saudi Arabia During COVID-19. *Front. Sustain. Cities* **2022**, *4*, 97. [CrossRef]
65. Alswedani, S.; Katib, I.; Abozinadah, E.; Mehmood, R. Discovering Urban Governance Parameters for Online Learning in Saudi Arabia During COVID-19 Using Topic Modeling of Twitter Data. *Front. Sustain. Cities* **2022**, *4*, 751681. [CrossRef]
66. Alomari, E.; Katib, I.; Albeshri, A.; Mehmood, R. COVID-19: Detecting Government Pandemic Measures and Public Concerns from Twitter Arabic Data Using Distributed Machine Learning. *Int. J. Environ. Res. Public Health* **2021**, *18*, 282. [CrossRef]
67. Alomari, E.; Katib, I.; Mehmood, R. Iktishaf: A Big Data Road-Traffic Event Detection Tool Using Twitter and Spark Machine Learning. *Mob. Netw. Appl.* **2020**, *21*, 6993. [CrossRef]
68. Alomari, E.; Mehmood, R.; Katib, I. Sentiment Analysis of Arabic Tweets for Road Traffic Congestion and Event Detection. In *Smart Infrastructure and Applications: Foundations for Smarter Cities and Societies*; Springer International Publishing: Cham, Switzerland, 2020; pp. 37–54.
69. Alsulami, M.; Mehmood, R. Sentiment Analysis Model for Arabic Tweets to Detect Users' Opinions about Government Services in Saudi Arabia: Ministry of Education as a Case Study. In Proceedings of the Al Yamamah Information and Communication Technology Forum, Riyadh, Saudi Arabia, 4–5 March 2018; pp. 1–8.

70. Suma, S.; Mehmood, R.; Albeshri, A. Automatic Detection and Validation of Smart City Events Using HPC and Apache Spark Platforms. In *Smart Infrastructure and Applications: Foundations for Smarter Cities and Societies*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 55–78.
71. Suma, S.; Mehmood, R.; Albugami, N.; Katib, I.; Albeshri, A. Enabling Next Generation Logistics and Planning for Smarter Societies. In *Proceedings of the Procedia Computer Science, Madeira, Portugal, 16–19 May 2017*; Elsevier: Amsterdam, The Netherlands, 2017; Volume 109, pp. 1122–1127.
72. Ramanathan, V.; Meyyappan, T. Twitter Text Mining for Sentiment Analysis on People's Feedback about Oman Tourism. In Proceedings of the 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), Muscat, Oman, 15–16 January 2019; pp. 1–5. [CrossRef]
73. Feizollah, A.; Mostafa, M.M.; Sulaiman, A.; Zakaria, Z.; Firdaus, A. Exploring Halal Tourism Tweets on Social Media. *J. Big Data* **2021**, *8*, 72. [CrossRef]
74. Hasnat, M.M.; Hasan, S. Identifying Tourists and Analyzing Spatial Patterns of Their Destinations from Location-Based Social Media Data. *Transp. Res. Part C Emerg. Technol.* **2018**, *96*, 38–54. [CrossRef]
75. Obembe, D.; Kolade, O.; Obembe, F.; Owoseni, A.; Mafimisebi, O. Covid-19 and the Tourism Industry: An Early Stage Sentiment Analysis of the Impact of Social Media and Stakeholder Communication. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100040. [CrossRef]
76. Al sari, B.; Alkhaldi, R.; Alsaffar, D.; Alkhaldi, T.; Almaymuni, H.; Alnaim, N.; Alghamdi, N.; Olatunji, S.O. Sentiment Analysis for Cruises in Saudi Arabia on Social Media Platforms Using Machine Learning Algorithms. *J. Big Data* **2022**, *9*, 21. [CrossRef]
77. Alasmari, W.A.; Abdelhafez, H.A. Twitter Sentiment Analysis for Reviewing Tourist Destinations in Saudi Arabia Using Apache Spark and Machine Learning Algorithms. *J. Comput. Sci.* **2022**, *18*, 215–226. [CrossRef]
78. Al-Smadi, M.; Al-Ayyoub, M.; Jararweh, Y.; Qawasmeh, O. Enhancing Aspect-Based Sentiment Analysis of Arabic Hotels' Reviews Using Morphological, Syntactic and Semantic Features. *Inf. Process. Manag.* **2019**, *56*, 308–319. [CrossRef]
79. Sayed, A.A.; Elgeldawi, E.; Zaki, A.M.; Galal, A.R. Sentiment Analysis for Arabic Reviews Using Machine Learning Classification Algorithms. In Proceedings of the 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE), Aswan, Egypt, 8–9 February 2020; pp. 56–63. [CrossRef]
80. Al Omari, M.; Al-Hajj, M.; Hammami, N.; Sabra, A. Sentiment Classifier: Logistic Regression for Arabic Services' Reviews in Lebanon. In Proceedings of the 2019 International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 3–4 April 2019; pp. 1–5. [CrossRef]
81. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the NAACL-HLT 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; Volume 1, pp. 4171–4186. [CrossRef]
82. McInnes, L.; Healy, J.; Astels, S. Hdbscan: Hierarchical Density Based Clustering. *J. Open Source Softw.* **2017**, *2*, 205. [CrossRef]
83. Grootendorst, M. BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure. *arXiv* **2022**, arXiv:2203.05794. [CrossRef]
84. Histograms—Matplotlib 3.6.0 Documentation. Available online: https://matplotlib.org/stable/gallery/statistics/hist.html (accessed on 9 October 2022).
85. Murtagh, F.; Contreras, P. Algorithms for Hierarchical Clustering: An Overview, II. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2017**, *7*, e1219. [CrossRef]
86. Sievert, C.; Shirley, K.E. LDAvis: A Method for Visualizing and Interpreting Topics. In Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, Baltimore, MA, USA, 27 June 2014; pp. 63–70.
87. Seaborn.Heatmap—Seaborn 0.11.2 Documentation. Available online: https://seaborn.pydata.org/generated/seaborn.heatmap.html (accessed on 9 October 2022).
88. Li, B.; Han, L. Distance Weighted Cosine Similarity Measure for Text Classification. *Lect. Notes Comput. Sci.* **2013**, *8206*, 611–618. [CrossRef]
89. Topic Visualization—BERTopic. Available online: https://maartengr.github.io/BERTopic/getting_started/visualization/visualization.html (accessed on 22 October 2022).
90. Waskom, M. Seaborn: Statistical Data Visualization. *J. Open Source Softw.* **2021**, *6*, 3021. [CrossRef]
91. Plotly Low-Code Data App Development. Available online: https://plotly.com/ (accessed on 23 October 2022).
92. Caruso, M.C.; Giuliano, R.; Pompei, F.; Mazzenga, F. Mobility Management for Smart Sightseeing. In Proceedings of the 2017 International Conference of Electrical and Electronic Technologies for Automotive, Turin, Italy, 15–16 June 2017. [CrossRef]
93. Naramski, M.; Herman, K. The Development of Mobile Tourism in the Upper Silesian Metropolitan Area of Poland. *Sustainability* **2019**, *12*, 44. [CrossRef]
94. Kong, H.; Labi, S.; Teng, H.; Karimoddini, A.; Chen, S.; Wu, F.; Ma, W. Clustering Analysis of the Spatio-Temporal On-Street Parking Occupancy Data: A Case Study in Hong Kong. *Sustainability* **2022**, *14*, 7957. [CrossRef]
95. Wang, Z.; Ye, Y.; Li, H.; Li, J.; Liu, J. Research on the Application of the Internet in the Development of Green Tourism Economy in China's Rural. In Proceedings of the 2020 2nd International Conference on Economic Management and Model Engineering (ICEMME), Chongqing, China, 20–22 November 2020; pp. 45–48. [CrossRef]
96. Cheuk, S.; Atang, A.; Chiun, L.M.; Ramayah, T. Barriers to Digital Marketing Adoption at Remote Rural Tourism Destinations in Sarawak: An Exploratory Study. *Int. J. Eng. Technol.* **2018**, *7*, 86–90. [CrossRef]

97. Peña, A.I.P.; Jamilena, D.M.F.; Molina, M.Á.R. The Perceived Value of the Rural Tourism Stay and Its Effect on Rural Tourist Behaviour. *J. Sustain. Tour.* **2012**, *20*, 1045–1065. [CrossRef]
98. Xue, L.; Kerstetter, D. Rural Tourism and Livelihood Change: An Emic Perspective. *J. Hosp. Tour. Res.* **2018**, *43*, 416–437. [CrossRef]
99. Martínez-Graña, A.M.; Serrano, L.; González-Delgado, J.A.; Dabrio, C.J.; Legoinha, P. Sustainable Geotourism Using Digital Technologies along a Rural Georoute in Monsagro (Salamanca, Spain). *Int. J. Digit. Earth* **2017**, *10*, 121–138. [CrossRef]
100. Van Huylenbroeck, G.; Vanslembrouck, I.; Calus, M.; van de Velde, L. Synergies between Farming and Rural Tourism: Evidence from Flanders. *EuroChoices* **2006**, *5*, 14–21. [CrossRef]
101. Rudiastuti, A.W.; Munawaroh; Setyawan, I.E.; Pramono, G.H. Coastal Management Strategy for Small Island: Ecotourism Potency Development in Karimata Island, West Kalimantan. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *148*, 12013. [CrossRef]
102. Eshun Tourism Programme, G.; Kumasi, P.; Ghana, R.; Tichaawa, T.M. Odame Appiah DGRD, D. Towards a Sustainable Coastal Tourism Development in Ghana. *Afr. J. Hosp. Tour. Leis.* **2019**, *8*, 1–18.
103. Phillips, M.R.; Jones, A.L. Erosion and Tourism Infrastructure in the Coastal Zone: Problems, Consequences and Management. *Tour. Manag.* **2006**, *27*, 517–524. [CrossRef]
104. Solis-Radilla, M.M.; Carvache-Franco, M.; Carvache-Franco, O.; Carvache-Franco, W. Motivations as Predictive Variables of Satisfaction and Loyalty in Coastal and Marine Destinations: A Study in Acapulco, Mexico. *Int. J. Tour. Cities* **2021**, *7*, 767–782. [CrossRef]
105. Zaman, M.B.; Djatmiko, E.B.; Nugroho, S.; Murdjito; Busse, W. Development of Safety for Marine Transportation in the Maratua Island. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *649*, 12066. [CrossRef]
106. Akla, C.M.N.; Suharyanto; Rachmad, B.; Firdaus, R.; Kurnianda, V. Fisheries Management Status in the Marine Protected Area of Eastern Coast of Weh Island, Sabang, Indonesia. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *674*, 12048. [CrossRef]
107. Emphandhu, D.; Polpiwat, U. Strategic Policy Options for Enhancement of National Park and Local Community Linkage in Tourism Management of Thai National Parks. *Kasetsart J. Soc. Sci.* **2006**, *27*, 347–362.
108. Rogowski, M. Monitoring System of Tourist Traffic (MSTT) for Tourists Monitoring in Mid-Mountain National Park, SW Poland. *J. Mt. Sci.* **2020**, *17*, 2035–2047. [CrossRef]
109. Liu, T.M. The Influence of Climate Change on Tourism Demand in Taiwan National Parks. *Tour. Manag. Perspect.* **2016**, *20*, 269–275. [CrossRef]
110. Kariyawasam, S.; Wilson, C.; Rathnayaka, L.I.M.; Sooriyagoda, K.G.; Managi, S. Conservation versus Socio-Economic Sustainability: A Case Study of the Udawalawe National Park, Sri Lanka. *Environ. Dev.* **2020**, *35*, 100517. [CrossRef]
111. Puustinen, J.; Pouta, E.; Neuvonena, M.; Sievänen, T. Visits to National Parks and the Provision of Natural and Man-Made Recreation and Tourism Resources. *J. Ecotourism* **2009**, *8*, 18–31. [CrossRef]
112. Hughes, K. Designing Post-Visit Action Resources for Families Visiting Wildlife Tourism Sites. *Visit. Stud.* **2011**, *14*, 66–83. [CrossRef]
113. Bueddefeld, J.N.H.; Van Winkle, C.M. Exploring the Effect of Zoo Post-Visit Action Resources on Sustainable Behavior Change. *J. Sustain. Tour.* **2017**, *25*, 1205–1221. [CrossRef]
114. Wright, D.W.M. Cloning Animals for Tourism in the Year 2070. *Futures* **2018**, *95*, 58–75. [CrossRef]
115. Geldenhuys, L. Determining the Market for Marine Wildlife Tourism in South Africa. *African J. Hosp. Tour. Leis.* **2019**, *8*, 1–8.
116. Shang, Z.; Luo, J.M. Topic Modelling for Wildlife Tourism Online Reviews: Analysis of Quality Factors. *Curr. Issues Tour.* **2022**, 1–15. [CrossRef]
117. Flower, E.K.; Burns, G.L.; Jones, D.N.; McBroom, J. Does the Experience Make a Difference? Comparing Tourist Attitudes Pre- and Post-Visit towards the Elephant Tourism Industry. *Ann. Tour. Res. Empir. Insights* **2021**, *2*, 100025. [CrossRef]
118. Carvache-Franco, M.; Carvache-Franco, W.; Carvache-Franco, O.; Alvarez-Risco, A.; Orden-Mejía, M.; Recalde-Lino, X. Designing an Adventure Tourism Package from the Preferences of the Visitors. *J. Environ. Manag. Tour.* **2022**, *XIII*, 305–312. [CrossRef] [PubMed]
119. Rosenberg, A.; Lynch, P.M.; Radmann, A. Sustainability Comes to Life. Nature-Based Adventure Tourism in Norway. *Front. Sport. Act. Living* **2021**, *3*, 154. [CrossRef] [PubMed]
120. Cordova-Buiza, F.; Calderon-Sanchez, J.; Chuzon-Canicela, L. Tourist Satisfaction Level: Evaluation in a Peruvian Adventure Tour Operator. *IBIMA Bus. Rev.* **2022**, *2022*, 773904. [CrossRef]
121. Senthilkumaran, P.; Pratim, S.P. Intervention Strategies to Mitigate Risk in Adventure Tourism: A Haddon Matrix Perspective. *Disaster Adv.* **2017**, *10*, 21–25.
122. Belias, D. Adventure Tourism-Examining Cases of How It Can Contribute on Rural Development in Greece. In Proceedings of the Springer Proceedings in Business and Economics; Technological Educational Institute of Thessaly, Larissa, Greece, 29 May 2019; pp. 903–910.
123. Dimopoulos, D.; Queiros, D.; van Zyl, C. Sinking Deeper: The Most Significant Risks Impacting the Dive Tourism Industry in the East African Marine Ecoregion. *Ocean Coast. Manag.* **2019**, *181*, 104897. [CrossRef]
124. Bideci, C.; Cater, C. In Search of Underwater Atmosphere: A New Diving World on Artificial Reefs. *Adv. Cult. Tour. Hosp. Res.* **2019**, *16*, 245–257.
125. Andy, L.; Lee, R.Y.; Tzeng, G.H. Characteristics of Professional Scuba Dive Guides. *Tour. Mar. Environ.* **2014**, *10*, 85–100. [CrossRef]

126. Tsilimigkas, G.; Rempis, N. Spatial Planning Framework, a Challenge for Marine Tourism Development: Location of Diving Parks on Rhodes Island, Greece. *Environ. Dev. Sustain.* **2021**, *23*, 15240–15265. [CrossRef]
127. Sorice, M.G.; Oh, C.-O.; Ditton, R.B. Managing Scuba Divers to Meet Ecological Goals for Coral Reef Conservation. *AMBIO A J. Hum. Environ.* **2007**, *36*, 316–322. [CrossRef]
128. Correia, A.I.; Nunes, A.; Silva, G.; Fernandes, P.O.; Moreira, J.; Soares, L.A. Sport Tourism Event and Perceived Economic Impacts: The Case of World Bodyboard Championship 2018, Viana Do Castelo, Portugal. *Smart Innov. Syst. Technol.* **2021**, *208*, 223–233. [CrossRef]
129. Lee, C.K.; Taylor, T. Critical Reflections on the Economic Impact Assessment of a Mega-Event: The Case of 2002 FIFA World Cup. *Tour. Manag.* **2005**, *26*, 595–603. [CrossRef]
130. Costa, A.; Fernandes, P.O.; Teixeira, J.P. The Importance of Cycling Sports in Regional Tourism—The Case of Volta a Portugal Em Bicicleta, Mondim de Basto Stage. *Smart Innov. Syst. Technol.* **2021**, *208*, 266–277. [CrossRef]
131. Achilleos, A.; Konstantinides, A.; Alexandrou, R.; Markides, C.; Zikouli, E.; Papadopoulos, G.A. A Web Platform and a Context Aware Recommender System for Active Sport Events. *Commun. Comput. Inf. Sci.* **2021**, *1404*, 183–197. [CrossRef]
132. Miyake, M.; Fujii, A.; Ohno, T.; Yoshikawa, M. Place-Based Services Platform That Enhances User Satisfaction from Sports Tourism to Daily Life. *FUJITSU Sci. Tech. J.* **2018**, *54*, 38–43.
133. Aravindhan, A.; Laxmikanth, G.; Shanmuga Priya, S.; Kamalraj, S. Medical Diagnosis during Space Tourism and Future Mars Colonization. In Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 6–7 March 2020; pp. 557–559. [CrossRef]
134. Saputra, J.; Harahap, M.A.K.; Sirojuzilam, S.; Harahap, R.H.; Marpaung, B.O.Y. The Effect of Open Tourism Space on Regional Development Through Supply Chain Management in Panatapan Danau Toba Area Simalungun, Indonesia. *Int. J. Supply Chain Manag.* **2019**, *8*, 994–1007.
135. Bukley, A.; Frize, R. Space Tourism: Risks and Solutions. In *Commercial Space Exploration: Ethics, Policy and Governance*; Routledge: London, UK, 2016; pp. 107–119.
136. Ganesha, K.S.; Sinnoor, G.B. Sentimental Analysis of Space Tourism: Evidence from Twitter. *New Space* **2021**, *9*, 148–150. [CrossRef]
137. Chang, E.Y.-W.; Chern, R.J.-S. Iaf symposium on commercial spaceflight safety issues (D6) enabling Safe Commercial Spaceflight: Vehicles and Spaceports (3) a preliminary study on the price model from aviation to suborbital to orbital space tourism. In Proceedings of the 69th International Astronautical Congress 2018, Bremen, Germany, 1–5 October 2018.
138. Toivonen, A. Sustainability Dimensions in Space Tourism: The Case of Finland. *J. Sustain. Tour.* **2020**, *30*, 2223–2239. [CrossRef]
139. Kadri, M.; Khalloufi, H.; Azough, A. V-Museum: A Virtual Museum Based on Augmented and Virtual Realities for Cultural Heritage Mediation. In Proceedings of the 2020 International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 9–11 June 2020. [CrossRef]
140. Liu, Y.; Liu, H.; Chris, C.; Chen, R. An Exploration of Digital Tourism Design Based on Virtual Reality. *Int. J. Simul. Syst. Sci. Technol.* **2016**, *17*, 1–15. [CrossRef]
141. Phithak, T.; Kamollimsakul, S. Korat Historical Explorer: The Augmented Reality Mobile Application to Promote Historical Tourism in Korat. *ACM Int. Conf. Proc. Ser.* **2020**, *17*, 283–289. [CrossRef]
142. Casillo, M.; Santo, M.D.; Lombardi, M.; Mosca, R.; Santaniello, D.; Valentino, C. Recommender Systems and Digital Storytelling to Enhance Tourism Experience in Cultural Heritage Sites. In Proceedings of the 2021 IEEE International Conference on Smart Computing (SMARTCOMP), Irvine, CA, USA, 23–27 August 2021; pp. 323–328. [CrossRef]
143. Sapio, F.; Ferro, L.S.; Mecella, M. Gaeta: The Great Adventure—A Cultural Heritage Game about the History of Gaeta. *Commun. Comput. Inf. Sci.* **2021**, *1421*, 179–187. [CrossRef]
144. Garlandini, A. Museums and Heritage in the Digital Age. The Challenge of Cultural Change and Technological Innovation. *SCIRES-IT Sci. Res. Inf. Technol.* **2021**, *11*, 11–18. [CrossRef]
145. Carneiro, M.J.; Eusébio, C.; Pelicano, M. An Expenditure Patterns Segmentation of the Music Festivals' Market. *Int. J. Sustain. Dev.* **2011**, *14*, 290–308. [CrossRef]
146. Lopes De Almeida, A.; Vieira, E.; Borges, A.P. Audience Segmentation and Communicating towards a Music Festival—The NOS Primavera Sound. Available online: https://www.proquest.com/openview/f91c607de35f4e5fe3277066adf9c1ff/1?pq-origsite=gscholar&cbl=4451210 (accessed on 1 February 2023).
147. Montoro-Pons, J.D.; Cuadrado-García, M. Analyzing Online Search Patterns of Music Festival Tourists. *Tour. Econ.* **2020**, *27*, 1276–1300. [CrossRef]
148. Tan, K.L.; Sim, A.K.S.; Chai, D.; Beck, L. Participant Well-Being and Local Festivals: The Case of the Miri Country Music Festival, Malaysia. *Int. J. Event Festiv. Manag.* **2020**, *11*, 433–451. [CrossRef]
149. Moisescu, O.I.; Gică, O.A.; Coroș, M.M.; Yallop, A.C. The UNTOLD Story: Event Tourism's Negative Impact on Residents' Community Life and Well-Being. *Worldw. Hosp. Tour. Themes* **2019**, *11*, 492–505. [CrossRef]
150. Wang, W.; Chen, J.S.; Huang, K. Religious Tourist Motivation in Buddhist Mountain: The Case from China. *Asia Pac. J. Tour. Res.* **2016**, *21*, 57–72. [CrossRef]
151. Al-Makhadmah, I.M. The Role of Virtual Museum in Promoting Religious Tourism in Jordan. *Geoj. Tour. Geosites* **2020**, *28*, 268–274. [CrossRef]
152. Albayrak, T.; Herstein, R.; Caber, M.; Drori, N.; Bideci, M.; Berger, R. Exploring Religious Tourist Experiences in Jerusalem: The Intersection of Abrahamic Religions. *Tour. Manag.* **2018**, *69*, 285–296. [CrossRef]

153. Marimin Cultural Heritage as a Tourist Destination: A Focus on Surakarta Kasunanan Palace in Indonesia. *J. Environ. Manag. Tour.* **2016**, *7*, 723–732. [CrossRef]
154. Ramos, C.M.Q.; Henriques, C.H.N.; Lanquar, R. Augmented Reality for Smart Tourism in Religious Heritage Itineraries: Tourism Experiences in the Technological Age. Available online: https://www.igi-global.com/chapter/augmented-reality-for-smart-tourism-in-religious-heritage-itineraries/158874 (accessed on 1 February 2023).
155. Kulshrestha, T.; Niyogi, R.; Misra, M.; Patel, D. Smart Pilgrim: A Mobile-Sensor-Cloud Based System to Safeguard Pilgrims through Smart Environment. In Proceedings of the 2017 Tenth International Conference on Contemporary Computing (IC3), Noida, India, 10–12 August 2017; pp. 1–6. [CrossRef]
156. Romanelli, M.; Gazzola, P.; Grechi, D.; Pollice, F. Towards a Sustainability-Oriented Religious Tourism. *Syst. Res. Behav. Sci.* **2021**, *38*, 386–396. [CrossRef]
157. Yen, H.P.; Chang, J.W.; Ho, K.C.; Hung, H.K. Foreign Muslim Workers' Perspectives of the Basic Needs of Muslim-Friendly Tourist Services. *Secur. Commun. Netw.* **2022**, *2022*, 1676697. [CrossRef]
158. Yahaya, M.Z.; Samsudin, M.A.; Kashim, M.I.A.M. An Analysis of Muslim Friendly Hotel Standards in Malaysia According to the Maqasid Syariah Perspective. *Int. J. Islam. Thought* **2020**, *18*, 43–53. [CrossRef]
159. Dabphet, S. Managing Islamic Attributes through the Satisfaction of Muslim Tourists in a Non-Muslim Country. *Int. J. Tour. Cities* **2021**, *7*, 237–254. [CrossRef]
160. Rahman, M.K.; Rana, M.S.; Ismail, M.N.; Muhammad, M.Z.; Hoque, M.N.; Jalil, M.A. Does the Perception of Halal Tourism Destination Matter for Non-Muslim Tourists' WOM? The Moderating Role of Religious Faith. *Int. J. Tour. Cities* **2022**, *8*, 478–496. [CrossRef]
161. Aji, H.M.; Muslichah, I.; Seftyono, C. The Determinants of Muslim Travellers' Intention to Visit Non-Islamic Countries: A Halal Tourism Implication. *J. Islam. Mark.* **2021**, *12*, 1553–1576. [CrossRef]
162. Sthapit, E.; Björk, P.; Piramanayagam, S. Motivational, Emotional and Memorable Dimensions of Non-Muslim Tourists' Halal Food Experiences. *J. Islam. Mark.* **2021**, *14*, 23–42. [CrossRef]
163. Jangra, P.; Gupta, M. Internet of Medical Things (IoMT) Based Framework for Smart Healthcare Tourism Sector. *Int. J. Hosp. Tour. Syst.* **2021**, *14*, 173–178.
164. Cham, T.H.; Lim, Y.M.; Sigala, M. Marketing and Social Influences, Hospital Branding, and Medical Tourists' Behavioural Intention: Before- and after-Service Consumption Perspective. *Int. J. Tour. Res.* **2022**, *24*, 140–157. [CrossRef]
165. Mahmud, M.S.; Lima, R.P.; Rahman, M.M.; Rahman, S. Does Healthcare Service Quality Affect Outbound Medical Tourists' Satisfaction and Loyalty? Experience from a Developing Country. *Int. J. Pharm. Healthc. Mark.* **2021**, *15*, 429–450. [CrossRef]
166. Prajitmutita, L.M.; Perényi, Á.; Prentice, C. Quality, Value?—Insights into Medical Tourists' Attitudes and Behaviors. *J. Retail. Consum. Serv.* **2016**, *31*, 207–216. [CrossRef]
167. Marković, S.; Lončarić, D.; hospitality, D.L.-T. Undefined Service Quality and Customer Satisfaction in the Health Care Industry-towards Health Tourism Market. *Tour. Hosp. Manag.* **2014**, *20*, 155–170. [CrossRef]
168. Smith, M. The Cost of Volunteering: Consequences of Voluntourism. Anthropology Senior Thesis, University of Pennsylvania, Philadelphia, PA, USA, 2015.
169. Petrzelka, P.; Mannon, S.E. Keepin' This Little Town Going: Gender and Volunteerism in Rural America. *Gend. Soc.* **2016**, *20*, 236–258. [CrossRef]
170. Sin, H.L. Volunteer tourism—"Involve me and i will learn"? *Ann. Tour. Res.* **2009**, *36*, 480–501. [CrossRef]
171. Chua, B.L.; Meng, B.; Ryu, H.B.; Han, H. Participate in Volunteer Tourism Again? Effect of Volunteering Value on Temporal Re-Participation Intention. *J. Hosp. Tour. Manag.* **2021**, *46*, 193–204. [CrossRef]
172. Jæger, K.; Olsen, K. On Commodification: Volunteer Experiences in Festivals†. *J. Tour. Cult. Chang.* **2016**, *15*, 407–421. [CrossRef]
173. Hallmann, K.; Zehrer, A. How Do Perceived Benefits and Costs Predict Volunteers' Satisfaction? *Volunt. Int. J. Volunt. Nonprofit Organ.* **2015**, *27*, 746–767. [CrossRef]
174. Amissah, E.F.; Opoku Mensah, A.; Mensah, I.; Gamor, E. Students' Perceptions of Careers in Ghana's Hospitality and Tourism Industry. *J. Hosp. Tour. Educ.* **2020**, *32*, 1–13. [CrossRef]
175. Lee, B.-W. Investigation on Current ETP Courses Implement in Taiwan's Universities. In Proceedings of the 2017 International Conference on Applied System Innovation (ICASI), Sapporo, Japan, 13–17 May 2017; pp. 981–984.
176. Didaskalou, E. Conceptual Modeling as an Approach to Designing a Sustainable Tourism Development Learning Programme. *WIT Trans. Ecol. Environ.* **2018**, *227*, 269–277. [CrossRef]
177. Ryndach, M.A.; Sheresheva, M. Open Education Model in Tourism. 2021. Available online: https://www.semanticscholar.org/paper/Open-Education-Model-in-Tourism-Ryndach-Sheresheva/c76428aeaff2803c1b777d434500900b162545bc (accessed on 1 February 2023).
178. Haarhoff, R.; Hass, A. Exploring the Link between Pleasant Tourist Experiences and Linguistic Competence: A Case Study of the Gariep Dam, Free State. *Afr. J. Hosp. Tour. Leis.* **2019**, *8*.
179. Bani-Melhem, S.; Zeffane, R.; Albaity, M. Determinants of Employees' Innovative Behavior. *Int. J. Contemp. Hosp. Manag.* **2018**, *30*, 1601–1620. [CrossRef]
180. Moradi, L.; Mohamed, I.; Yahya, Y. The Effect of Organizational Commitment and E-Training on E-Tourism Job Performance. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2018**, *8*, 2286–2293. [CrossRef]

181. Tsai, C.W. Leadership Style and Employee's Job Satisfaction in International Tourist Hotels. *Adv. Cult. Tour. Hosp. Res.* **2008**, *2*, 293–332. [CrossRef]
182. Choi, H.; Lee, J.Y.; Choi, Y.; Juan, Y.; Lee, C.K. How to Enhance Smart Work Effectiveness as a Sustainable HRM Practice in the Tourism Industry. *Sustainability* **2022**, *14*, 2218. [CrossRef]
183. Natércia Gonçalves Costa, V. Tourism and Air Transport—An Economic Evaluation of the Oporto Airport Expansion Project. *Tour. Manag. Stud.* **2020**, *16*, 35–42. [CrossRef]
184. Çetin, G.; Akova, O.; Gursoy, D.; Kaya, F.K.F. Impact of Direct Flights on Tourist Volume: Case of Turkish Airlines. *J. Tour.* **2016**, *2*, 36–50. [CrossRef]
185. Hassan, T.H.; Salem, A.E.; Choi, H.-M.; Sung, H.-J. The Importance of Safety and Security Measures at Sharm El Sheikh Airport and Their Impact on Travel Decisions after Restarting Aviation during the COVID-19 Outbreak. *Sustainability* **2021**, *13*, 5216. [CrossRef]
186. Miočić, B.K.; Vidić, G.; Klarin, T. Comparative Analysis of Tourist Satisfaction and Online Booking Services Usage for Incoming Tourists in Zadar County. In Proceedings of the 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 26–30 May 2014; pp. 1544–1549. [CrossRef]
187. Moreno-Izquierdo, L.; Ramón-Rodríguez, A.; Perles Ribes, J. The Impact of the Internet on the Pricing Strategies of the European Low Cost Airlines. *Eur. J. Oper. Res.* **2015**, *246*, 651–660. [CrossRef]
188. Çakar, K.; Aykol, Ş. Understanding Travellers' Reactions to Robotic Services: A Multiple Case Study Approach of Robotic Hotels. *J. Hosp. Tour. Technol.* **2021**, *12*, 155–174. [CrossRef]
189. Ho, J.-L.; Chen, K.-Y.; Wang, L.-H.; Yeh, S.-S.; Huan, T.-C. Exploring the Impact of Social Media Platform Image on Hotel Customers' Visit Intention. *Int. J. Contemp. Hosp. Manag.* 2022, *ahead-of-print*. [CrossRef]
190. Wang, X.; Xu, J. Deterrence and Leadership Factors: Which Are Important for Information Security Policy Compliance in the Hotel Industry. *Tour. Manag.* **2021**, *84*, 104282. [CrossRef]
191. Abo Murad, M.; Al-Kharabsheh, A.; Al-Kharabsheh, A. Crisis Management Strategies in Jordanian Hotel Industry. *J. Environ. Manag. Tour.* **2021**, *12*, 578–587. [CrossRef]
192. Martínez, P.; Herrero, Á.; Gómez-López, R. Corporate Images and Customer Behavioral Intentions in an Environmentally Certified Context: Promoting Environmental Sustainability in the Hospitality Industry. *Corp. Soc. Responsib. Environ. Manag.* **2019**, *26*, 1382–1391. [CrossRef]
193. Elmont, S. Tourism and Food Service: Two Sides of the Same Coin. *Cornell Hosp. Q.* **2016**, *36*, 57–63. [CrossRef]
194. Ivanov, S.; Webster, C. Restaurants and Robots: Public Preferences for Robot Food and Beverage Services. *J. Tour. Futur.* 2022, *ahead-of-print*. [CrossRef]
195. Adriatico, R.L.; Razalan, A.M.A.; Pagbilao, C.M.V.; Afalla, B.T.; Cruz, L.M. Dela Service Quality and Customer Satisfaction in Dining Restaurants: Inputs for Tourism and Hospitality Curriculum Enhancement. *Acad. J. Interdiscip. Stud.* **2022**, *11*, 30. [CrossRef]
196. Huang, C.-H. Sustainable Strategies of Restaurant Food Surplus Platform as a Framework for Responsible Tourism in the Sharing Economy. *E-Rev. Tour. Res.* **2020**, *17*, 633–642.
197. Lee, Y.J.; Pennington-Gray, L.; Kim, J. Does Location Matter? Exploring the Spatial Patterns of Food Safety in a Tourism Destination. *Tour. Manag.* **2019**, *71*, 18–33. [CrossRef]
198. Kotlyarov, N.N.; Chuvakhina, L.G.; Terskaya, G.A. Tourism as an Important Tool in the International Strategy for Sustainable Development. In *Financial and Economic Tools Used in the World Hospitality Industry*; CRC Press: Boca Raton, FL, USA, 2018; pp. 121–126.
199. Hussein, S.H.; Kusairi, S.; Ismail, F. Modelling the Demand for Educational Tourism: Do Dynamic Effect, University Quality and Competitor Countries Play a Role? *J. Tour. Futur.* 2022, *ahead-of-print*. [CrossRef]
200. Al-Hammadi, A.; Al-Shami, S.A.; Al-Hammadi, A.; Rashid, N. Halal Tourism Destination in UAE: The Opportunities, Threats and Future Research. *Undefined* **2019**, *8*, 788–793. [CrossRef]
201. Din, B.H.; Habibullah, M.S.; Tan, S.H. The Effects of World Heritage Sites and Governance on Tourist Arrivals: Worldwide Evidence. *Int. J. Econ. Manag.* **2017**, *11*, 437–448.
202. Zeng, B.; Ryan, C.; Cui, X.; Chen, H. Tourism-Generated Income Distribution in a Poor Rural Community: A Case Study from Shaanxi, China. *J. China Tour. Res.* **2014**, *11*, 85–104. [CrossRef]
203. Sati, V.P. *Development of Eco-Tourism and Hydroelectricity*; Environmental Science and Engineering; Springer: Cham, Switzerland, 2014; pp. 91–101. [CrossRef]
204. Cole, S. *Tourism, Culture and Development: Hopes, Dreams and Realities in East Indonesia*; Blue Ridge Summit: Franklin County, PA, USA; Channel View Publications: Bristol, UK, 2007.
205. Lu, J.; Deng, A.; Ye, P. Rural Destination Revitalization in Shennongjia Forestry District: A Tourism Poverty Alleviation Perspective. In Proceedings of the ICEB 2020 Proceedings, Hong Kong, China, 5–8 December 2020; pp. 1–8.
206. Holden, A.; Sonne, J.; Novelli, M. Tourism and Poverty Reduction: An Interpretation by the Poor of Elmina, Ghana. *Ourism Plan. Dev.* **2011**, *8*, 317–334. [CrossRef]
207. Önder, I.; Gunter, U.; Scharl, A. Forecasting Tourist Arrivals with the Help of Web Sentiment: A Mixed-Frequency Modeling Approach for Big Data. *Tour. Anal.* **2019**, *24*, 437–452. [CrossRef]

208. Dananjali, T.; Wijesinghe, S.; Ekanayake, J. Forecasting Weekly Rainfall Using Data Mining Technologies. In Proceedings of the 2020 From Innovation to Impact, FITI 2020, Colombo, Sri Lanka, 15 December 2020.
209. Wu, S. Forecasting Museum Visitor Behaviors Using Deep Learning. In Proceedings of the 2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), Taiyuan, China, 8–10 November 2019; pp. 186–190. [CrossRef]
210. Keliwar, S.; Putra, A.B.W.; Hammad, J. Haviluddin Modeling of Time Series Data for Forecasting the Number of Foreign Tourists in East Kalimantan Using Fuzzy Inference System Based on ARX Model. *Undefined* **2018**, *7*, 104–107. [CrossRef]
211. Gazioğlu, C.; Burak, S.; Alpar, B.; Türkerd, A.; Barutc, I.F. Foreseeable Impacts of Sea Level Rise on the Southern Coast of the Marmara Sea (Turkey). *Water Policy* **2010**, *12*, 932–943. [CrossRef]
212. Hůnová, I. Surface Ozone Measurements in Mountain Forested Areas Using Diffusive Samplers. *WIT Trans. Ecol. Environ.* **2006**, *86*, 585–594. [CrossRef]
213. Chunjou, M.; Pang, S.F.H. An Exploratory Study of Corporate Social Responsibility of Travel Agency Websites and Consumers' Low Carbon Travel Intention. In Proceedings of the 2013 Seventh International Conference on Complex, Intelligent, and Software Intensive Systems, Taichung, Taiwan, 3–5 July 2013; pp. 661–666. [CrossRef]
214. Liu, F.; Sun, J.; Liu, M.; Yang, J.; Gui, G. Generalized Flight Delay Prediction Method Using Gradient Boosting Decision Tree. In Proceedings of the 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium, 25–28 May 2020. [CrossRef]
215. Erdoğan, S.; Gedikli, A.; Cevik, E.I.; Erdoğan, F. Eco-Friendly Technologies, International Tourism and Carbon Emissions: Evidence from the Most Visited Countries. *Technol. Forecast. Soc. Chang.* **2022**, *180*, 121705. [CrossRef]
216. Abdull, N.; Yoneda, M.; Shimada, Y. Traffic Characteristics and Pollutant Emission from Road Transport in Urban Area. *Air Qual. Atmos. Health* **2020**, *13*, 731–738. [CrossRef]
217. Pagoni, I.; Psaraki, V. A Tool for Calculating Aircraft Emissions and Its Application to Greek Airspace. *Transp. Plan. Technol.* **2014**, *37*, 138–153. [CrossRef]
218. Liu, H.; Jiang, Y.; Zhu, H.; Chen, Y.; Lyu, W.; Luo, W.; Yao, W. Analysis of Water Resource Management in Tourism in China Using a Coupling Degree Model. *Water Policy* **2021**, *23*, 765–782. [CrossRef]
219. Khan, R.H.R.; Somphong, W.; Naklang, S.; Boondesh, N.; Layanan, P. Impact of Tourism Activities on Water Pollution and Tourist Attractions in Koh Larn Beach Pattaya, Thailand. *Int. J. Adv. Sci. Technol.* **2020**, *29*, 1517–1522.
220. Diaz Perez, F.J.; Pino Otin, M.R.; Mouhaffel, A.G.; Martin, R.D.; Chinarro, D. Energy and Water Consumption and Carbon Footprint in Tourist Pools Supplied by Desalination Plants: Case Study, the Canary Islands. *IEEE Access* **2018**, *6*, 11727–11737. [CrossRef]
221. Mishra, P.; Panda, U.S.; Pradhan, U.; Kumar, C.S.; Naik, S.; Begum, M.; Ishwarya, J. Coastal Water Quality Monitoring and Modelling off Chennai City. *Procedia Eng.* **2015**, *116*, 955–962. [CrossRef]
222. Salgot, M.; Oron, G.; Cirelli, G.L.; Dalezios, N.R.; Díaz, A.; Angelakis, A.N. Criteria for Wastewater Treatment and Reuse under Water Scarcity. In *Handbook of Drought and Water Scarcity*; CRC Press: Boca Raton, FL, USA, 2017; pp. 263–282. [CrossRef]
223. Lan, T.; Yang, Y.; Shao, Y.; Luo, M.; Zhong, F. The Synergistic Effect of Natural Disaster Frequency and Severity on Inbound Tourist Flows from the Annual Perspective. *Tour. Manag. Perspect.* **2021**, *39*, 100832. [CrossRef]
224. Gani, A.; Singh, R.; Najar, A.H. Rebuilding Tourist Destinations from Crisis: A Comparative Study of Jammu and Kashmir and Assam, India. *Worldw. Hosp. Tour. Themes* **2021**, *13*, 437–454. [CrossRef]
225. Nguyen, C.P. Last Chance to Travel or Safety First? The Influence of Exposure to Natural Hazards and Coping Capacities on Tourism Consumption. *Tour. Econ.* **2022**, 13548166221077648. [CrossRef]
226. Uysal, M.; Toprak, A.S.; Polat, N. Photo Realistic 3D Modeling wth Uav: Gedik Ahmet Pasha Mosque in Afyonkarahisar. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, *W2*, 659–662. [CrossRef]
227. Kovači'c, S.K.; Mărgărint, M.C.; Ionce, R.; Da, Đ.¯; Miljkovi´c, M. What Are the Factors Affecting Tourist Behavior Based on the Perception of Risk? Romanian and Serbian Tourists' Perspective in the Aftermath of the Recent Floods and Wildfires in Greece. *Sustainability* **2020**, *12*, 6310. [CrossRef]
228. Triastutiningsih, C.; Rachmawati, R. Utilization of the Visiting Jogja Mobile Application as a Provider of Information Regarding Limitations of Tourism Activities during the COVID-19 Pandemic in the Special Region of Yogyakarta. *Lect. Notes Netw. Syst.* **2022**, *393*, 53–63. [CrossRef]
229. Ghosh, S. Modelling Inbound International Tourism Demand in Australia: Lessons from the Pandemics. *Int. J. Tour. Res.* **2021**, *24*, 71–81. [CrossRef]
230. Wang, J.; Ge, P.; Liu, Z. Using Denoised LSTM Network for Tourist Arrivals Prediction. In Proceedings of the 2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML), Chengdu, China, 16–18 July 2021; pp. 176–182. [CrossRef]
231. Polese, F.; Botti, A.; Monda, A. Value Co-Creation and Data-Driven Orientation: Reflections on Restaurant Management Practices during COVID-19 in Italy. *Transform. Gov. People Process. Policy* **2022**, *16*, 172–184. [CrossRef]
232. Agrawal, A. Sustainability of Airlines in India with Covid-19: Challenges Ahead and Possible Way-Outs. *J. Revenue Pricing Manag.* **2021**, *20*, 457–472. [CrossRef]
233. Pilgrimage during COVID-19: Impacts, Adaptations, and Recovery. 2022. Available online: https://ixtheo.de/Record/1801063346 (accessed on 1 February 2023).

234. Munawar, H.S.; Khan, S.I.; Ullah, F.; Kouzani, A.Z.; Parvez Mahmud, M.A. Effects of COVID-19 on the Australian Economy: Insights into the Mobility and Unemployment Rates in Education andtourism Sectors. *Sustainability* **2021**, *13*, 11300. [CrossRef]
235. König, M.; Winkler, A. COVID-19 and Economic Growth: Does Good Government Performance Pay Off? *Intereconomics* **2020**, *55*, 224–231. [CrossRef] [PubMed]
236. Bolderman, L. Locating Imagination: An Interdisciplinary Perspective on Literary, Film, and Music Tourism. *Tour. Anal.* **2015**, *20*, 333–339. [CrossRef]
237. Nieto-Ferrando, J.; Sánchez-Castillo, S.; Gómez-Morales, B. Audiovisual Fiction and Tourism Promotion: The Impact of Film and Television on the Image of Tourist Destinations and Contributions from Textual Analysis. *Profesional de la información* **2021**, *30*, 1699–2407. [CrossRef]
238. Iwashita, C. Roles of Films and Television Dramas in International Tourism: The Case of Japanese Tourists to the UK. *J. Travel Tour. Mark.* **2010**, *24*, 139–151. [CrossRef]
239. Garrison, S.; Wallace, C. Media Tourism and Its Role in Sustaining Scotland's Tourism Industry. *Sustainability* **2021**, *13*, 6305. [CrossRef]
240. Kay, A. Promoting Tourism Through Popular Music. *Tour. Cult. Commun.* **2006**, *6*, 209–213. [CrossRef]
241. García-Pablos, A.; Cuadros, M.; Linaza, M.T. Automatic Analysis of Textual Hotel Reviews. *Inf. Technol. Tour.* **2015**, *16*, 45–69. [CrossRef]
242. Yang, Y.; Lin, X. Tourism Network Comments Sentiment Analysis and Early Warning System Based on Ontology. *Lect. Notes Comput. Sci.* **2016**, *9772*, 863–870. [CrossRef]
243. Park, E.; Park, J.; Hu, M. Tourism Demand Forecasting with Online News Data Mining. *Ann. Tour. Res.* **2021**, *90*, 103273. [CrossRef]
244. Tang, L.; Zhang, C.; Li, T.; Li, L. A Novel BEMD-Based Method for Forecasting Tourist Volume with Search Engine Data. *Tour. Econ.* **2020**, *27*, 1015–1038. [CrossRef]
245. Thenmozhi, M.; Harshitha, S.; Gayathidevi, M.; Sreenija Reddy, C. A Framework for Tourist Recommendation System Exploiting Geo-Tagged Photos. In Proceedings of the 2016 10th International Conference on Intelligent Systems and Control (ISCO), Coimbatore, India, 7–8 January 2016. [CrossRef]
246. Hasnat, M.M.; Hasan, S. Understanding Tourist Destination Choices from Geo-Tagged Tweets. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 3391–3396.
247. Rubegni, E.; Gerardi, S.; Caporali, M. Mobile Applications for Helping Users to Keep Track of Their Travel Experience. *ACM Int. Conf. Proc. Ser.* **2007**, *250*, 311–312. [CrossRef]
248. Yang, S.Y.; Hsu, C.L. A Location-Based Services and Google Maps-Based Information Master System for Tour Guiding. *Comput. Electr. Eng.* **2016**, *54*, 87–105. [CrossRef]
249. Nasir, S.; Al-Qaraawi, S.; Croock, M. Design and Implementation a Network Mobile Application for Plants Shopping Center Using QR Code. *Int. J. Electr. Comput. Eng.* **2020**, *10*, 5940–5950. [CrossRef]
250. Huettermann, M.; Thimm, T.; Hannich, F.; Bild, C. Requirements for Future Digital Visitor Flow Management. *J. Tour. Futur.* **2019**, *5*, 241–258. [CrossRef]
251. Koeva, M.N. 3D Modelling and Interactive Web-Based Visualization of Cultural Heritage Objects. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. ISPRS Arch.* **2016**, *41*, 297–303. [CrossRef]
252. Singthongchai, J.; Naenudorn, E.; Kittidachanupap, N.; Khopolklang, N.; Niwattanakul, S. Virtual 3-D Animation for Tourism. In Proceedings of the 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE), Zhangjiajie, China, 25–27 May 2012; Volume 2, pp. 102–106. [CrossRef]
253. Feld, N.; Weyers, B. Mixed Reality in Asymmetric Collaborative Environments: A Research Prototype for Virtual City Tours. In Proceedings of the 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Lisbon, Portugal, 27 March–1 April 2021; pp. 250–256. [CrossRef]
254. García-Crespo, Á.; González-Carrasco, I.; López-Cuadrado, J.L.; Villanueva, D.; González, Á. CESARSC: Framework for Creating Cultural Entertainment Systems with Augmented Reality in Smart Cities. *Comput. Sci. Inf. Syst.* **2016**, *13*, 395–425. [CrossRef]
255. Lovasz-Bukvova, H.; Hölzl, M.; Kormann-Hainzl, G.; Moser, T.; Zigart, T.; Schlund, S. *Usability and Task Load of Applications in Augmented and Virtual Reality: How Applicable Are the Technologies in Corporate Settings?* Springer International Publishing: Cham, Switzerland, 2021; Volume 1442, ISBN 9783030855208.
256. Ainin, S.; Feizollah, A.; Anuar, N.B.; Abdullah, N.A. Sentiment Analyses of Multilingual Tweets on Halal Tourism. *Tour. Manag. Perspect.* **2020**, *34*, 100658. [CrossRef]
257. Yuan, H.; Xu, H.; Qian, Y.; Li, Y. Make Your Travel Smarter: Summarizing Urban Tourism Information from Massive Blog Data. *Int. J. Inf. Manage.* **2016**, *36*, 1306–1319. [CrossRef]
258. Silva, B.; Moro, S.; Marques, C. Sensing the Impact of COVID-19 Restrictions from Online Reviews: The Cases of London and Paris Unveiled Through Text Mining. *Smart Innov. Syst. Technol.* **2022**, *279*, 223–232. [CrossRef]
259. Chang, Y.C.; Ku, C.H.; Nguyen, D.D. Le Predicting Aspect-Based Sentiment Using Deep Learning and Information Visualization: The Impact of COVID-19 on the Airline Industry. *Inf. Manag.* **2022**, *59*, 103587. [CrossRef]
260. News, A. Saudi Arabia's Crown Prince Announces 'The Line'–A Zero Carbon City of 1 Million | Arab News Japan. Available online: https://www.arabnews.com/node/1790351/saudi-arabia (accessed on 1 February 2023).

261. Mehmood, R. 'Deep Journalism' Driven by AI Can Aid Better Government. *The Mandarin*, 1 October 2022. Available online: https://www.themandarin.com.au/201467-deep-journalism-driven-by-ai-can-aid-better-government/ (accessed on 1 February 2023).
262. Alaql, A.A.; Alqurashi, F.; Mehmood, R. Data-Driven Deep Journalism to Discover Age Dynamics in Multi-Generational Labour Markets from LinkedIn Media. *J. Media* **2023**, *4*, 120–145. [CrossRef]
263. Alswedani, S.; Mehmood, R.; Katib, I.; Altowaijri, S.M. Psychological Health and Drugs: Data-Driven Discovery of Causes, Treatments, Effects, and Abuses. *Preprints* **2023**, 2023010415. [CrossRef]
264. Pachura, P.; Nitkiewicz, T.; Matlovičová, K.; Matlovič, R. Identification of Intellectual Capital Performance Using Data Envelopment Analysis. *Adv. Spat. Sci.* **2018**, 115–130. [CrossRef]
265. Neumannová, M. Smart Districts: New Phenomenon in Sustainable Urban Development. Case Study of Špitálka in Brno, Czech Republic. *Folia Geogr.* **2022**, *64*, 27–48.

sustainability

Review

Crowdsourcing Public Engagement for Urban Planning in the Global South: Methods, Challenges and Suggestions for Future Research

El Bachir Diop [1,*], Jérôme Chenal [1,2], Stéphane Cédric Koumetio Tekouabou [1] and Rida Azmi [1]

[1] Center of Urban Systems, Mohammed VI Polytechnic University, No. 660, Hay Moulay Rachid, Benguérir 43150, Morocco
[2] Communauté d'Etudes pour l'Aménagement du Territoire (CEAT/EPFL), École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland
* Correspondence: elbachir.diop@um6p.ma

Abstract: Crowdsourcing could potentially have great benefits for the development of sustainable cities in the Global South (GS), where a growing population and rapid urbanization represent serious challenges for the years to come. However, to fulfill this potential, it is important to take into consideration the unique characteristics of the GS and the challenges associated with them. This study provides an overview of the crowdsourcing methods applied to public participation in urban planning in the GS, as well as the technological, administrative, academic, socio-economic, and cultural challenges that could affect their successful adoption. Some suggestions for both researchers and practitioners are also provided.

Keywords: crowdsourcing; Global South; urban planning; developing countries; public participation; big urban data

Citation: Diop, E.B.; Chenal, J.; Tekouabou, S.C.K.; Azmi, R. Crowdsourcing Public Engagement for Urban Planning in the Global South: Methods, Challenges and Suggestions for Future Research. *Sustainability* **2022**, *14*, 11461. https://doi.org/10.3390/su141811461

Academic Editors: Tan Yigitcanlar, Bo Xia, Tatiana Tucunduva Philippi Cortese and Jamile Sabatini Marques

Received: 2 August 2022
Accepted: 6 September 2022
Published: 13 September 2022

1. Introduction

Although the concept of crowdsourcing is fairly recent [1], the idea of engaging the public and non-experts in problem-solving and data collection has a long history in both research and practice. In 1936, the Japanese company Toyota (then Toyoda) organized a public contest for the design of its new logo [2]. In total, 27,000 designs were submitted, and the best logo was selected and used between 1936 and 1989. In the 1960s, public advocacy theory [3] (emphasized the importance of public participation in urban planning. Other concepts, such as citizen science, and Public Participation Geographic Information Systems (PPGIS), follow the same principle of engaging the public to participate in the design and implementation of solutions to various problems regardless of their level of expertise. With the upsurge of the Internet, researchers and practitioners had to rethink the ways in which public participation is carried out and re-assess the societal transformation that comes with it. This led to the emergence of crowdsourcing, "a web-based business model requiring voluntary open collaboration to develop innovative solutions" [1]. By tapping into a large and diverse pool of stakeholders through the Internet and Web 2.0 technologies, crowdsourcing, as a public participation method, has alleviated the spatial and temporal constraints that are associated with the aforementioned methods.

The term Global South (GS) has several definitions which have economic, geopolitical, and cultural implications. Economically speaking, the GS groups developing countries characterized by, among other indicators, medium and low human development index (HDI less than 0.8). Geographically speaking, most of the GS is in the southern hemisphere and regroups African, Southern and Central American, and Asian countries (with the exception of Japan, South Korea, and Singapore). Due to their limited resources, these countries struggle to develop plans that could effectively address the challenges faced by contemporary cities. For example, in the Asia-Pacific region, over 50% of the Sustainable Development Goals (SDGs) cannot be measured due to a lack of data [4]. Fraisl et al. [5,6] have demonstrated that crowdsourcing could help monitor SDG indicators. Thus, crowdsourcing could be very useful to these countries as it allows gathering useful data, which in turn could support better-informed policies. From a historical point of view, most of these countries are former European colonies. As such, the traditional urban planning method consisted mainly of copying strategies implemented in the former colonial power [7,8]. However, these strategies are rarely successful as they fail to take into account the unique challenges faced by developing countries [9]. With its participatory approach, crowdsourcing could provide a platform that taps into the citizens' local knowledge to identify the main challenges faced by cities in the GS. This, in turn, could help planners better define their priorities and implement policies that meet the needs of the local communities. Furthermore, this democratized planning process through public participation can lead to more transparency and greater citizens' acceptance of public decisions [10].

In the Global North, crowdsourcing has helped democratize the planning process, empower citizens, provide low-cost data for real-time planning, and helped mitigate the limitations of traditional data collection methods such as census data [11–15]. In America, Thiagarajan et al. [12] used low-cost, grassroots GPS tracking solutions to improve riders' transit experience (e.g., reduction of waiting time), while Griffin and Jiao [15] demonstrated that collecting data through crowdsourcing increased the inclusiveness of the participatory planning process from the perspective of geography and equity. In Australia, K. Hu et al. [14] developed a low-cost participatory sensing system (called HazeWatch) for urban air pollution monitoring which yields more accurate measurements than the existing government system. The HazeWatch system provides a better understanding of the health impact of air pollution in metropolitan areas. In Italy, MiraMap [13], a we-government platform, helps facilitate the collaboration between the public and the administration while promoting social inclusion, transparency, and accountability in smart city management. These examples in the Global North show the potential value crowdsourcing could have for the GS, which is characterized by limited resources, low or inexistent citizen participation, and a lack of transparency, accountability, and data-driven planning methods.

However, despite these possible advantages, the potential of crowdsourcing remains to be exploited in the GS. This is even more true in Africa, where most urban studies rely on qualitative analysis or traditional data collection methods (survey questionnaires) instead of quantitative methods that require abundant and reliable data [16]. This affects the reliability of the findings and limits the effectiveness of the policies that could be implemented from the existing literature. Furthermore, the existing reviews on crowdsourcing in urban studies mainly provide a global overview of the literature [17–22]. These studies provide a clear understanding of the methods and challenges associated with the use of crowdsourcing. However, the GS faces specific cultural, technological, political, and administrative challenges which could greatly impact the successful use of crowdsourcing in this part of the world. To fill this gap, this paper presents a review of the crowdsourcing research efforts conducted in the GS. More specifically, this study describes the crowdsourcing methods adopted in the GS as well as the main areas of application. The methods described focus on public engagement to support urban planning. Therefore, crowdsourcing in this context mainly consists of data collected and shared by the public through mobile devices (GPS tracking, crash reporting, environment monitoring, etc.) and/or local knowledge shared through collaborative websites (crime mapping, flood mapping, idea generation for smart city management, etc.). Furthermore, drawing from the descriptive statistics of the reviewed papers as well as the characteristics of the GS, the paper also discusses the challenges that could hinder the implementation of crowdsourcing. Finally, it suggests some solutions that could be useful to developing countries in general. This approach has some advantages, which could lead to significant contributions to the existing literature:

- By providing an overview of the main areas of research, we identify the domains where more research is needed in the future;

- Drawing lessons from countries that share the same historical, social, and economical experiences seems more logical than copying methods adopted in the developed world and could lead to more realistic solutions.

The remainder of this paper is organized as follows. The next section discusses the concept of crowdsourcing and adapts it to the context of this study. Section 3 outlines the review method and provides a summary of the reviewed literature. Sections 4 and 5 identify the main areas of application and methods, respectively. Section 6 discusses the challenges associated with the development of crowdsourcing methods in the GS and provides suggestions for future implementations. Section 7 concludes this study.

2. Crowdsourcing: Definitions

Since Howe [1], several studies have provided different definitions of crowdsourcing. These definitions are important as they provide a basis for what should be considered crowdsourcing and what should not. For example, some studies perceive YouTube and Wikipedia as crowdsourcing [23] while others do not [24].

In urban planning, concepts such as problem-solving, idea generation, and collaborative mapping are widely accepted as crowdsourcing [23,25–27], while data collection methods such as social media scraping and crowdsensing are subject for debate [25,28] Brabham [25] defines crowdsourcing as a top-down approach to solving planning problems. This definition includes approaches such as idea generation for smart city solutions [23,29] but excludes data collection methods such as crowdsensing, Public Participation Geographic Information Systems (PPGIS), social media, etc. Nakatsu et al. [28] argue for a broader definition that includes "geo-located data collection" (e.g., GPS tracking, a form of crowdsensing) but excludes social media. Their main argument for excluding social media was the absence of explicit outsourcing of a task *to* the crowd. Furthermore, although social media have been widely adopted as crowdsourced data, the method usually consists of extracting people's posts (social media scraping) through Application Programming Interfaces (APIs) without their consent. This could raise some ethical concerns as the people whose posts are extracted may not be willing to participate in data collection. Besides, Howe, who introduced the concept of crowdsourcing, also defined it as a voluntary process. Finally, Estellés-Arolas and González-Ladrón-De-Guevara [24] have provided a definition of crowdsourcing based on a thorough review of the existing literature. They found voluntary participation and a clearly defined task among the main criteria for crowdsourcing. Based on the aforementioned studies, the adoption of methods that do not necessarily require voluntary participation (such as social media scraping and crowdsensing) may be problematic. However, it would be too simplistic to discard all studies using social media or crowdsensing without exploring cases where the participation is voluntary and the task clearly defined. The next subsections address this issue in detail.

2.1. Social Media Data

Although most studies use social media scraping, there are specific cases in which the methods described meet the criteria we described above. These cases are:

- Voluntary participation in dedicated social media groups or pages. Dedicated social media pages can be open platforms for citizen engagement. In this case, the task could consist of submitting complaints (e.g., HarassMapEgypt, a Facebook page [30]), participating in e-governance or sharing citizen sensing data (e.g., pictures, videos, etc.), etc. (see Section 5.3).
- Studies using social media scraping as a primary data collection method and another crowdsourcing method (usually Open Street Map, OSM) as a secondary dataset. We believe such studies to be of importance as they demonstrate how crowdsourcing could complement other datasets.

2.2. Crowdsensing

Crowdsensing leverages the proliferation of low-cost sensing devices and citizen engagement for collecting and sharing data in different domains (environment monitoring, traffic management, waste management, etc.). Participation in crowdsensing can be voluntary or non-voluntary. For example, a crowdsensing application can combine sensing data (e.g., GPS data) with lobation-based service network datasets such as social media check-ins [31]. Thus, similar to social media, we will carefully identify the studies in which participation in crowdsensing is voluntary.

Web-based PPGIS has also been used to crowdsource data for urban planning [32]. Some Web-based PPGIS projects provide an online platform where participants can share local knowledge through open calls, which is consistent with the basic principles of crowdsourcing.

Therefore, in line with the arguments discussed above, we adopt a broader definition of crowdsourcing which covers voluntary crowdsensing, dedicated social media campaigns, and collaborative websites (web-based PPGIS, collaborative mapping, and idea generation).

3. Method and Descriptive Statistics of the Reviewed Papers

3.1. Literature Search

In this section, we adopt the PRISMA [33] method to search for the core literature used in this study. PRISMA (Figure 1) consists of the following steps: identification, screening, eligibility, and inclusion. We discuss each step below.

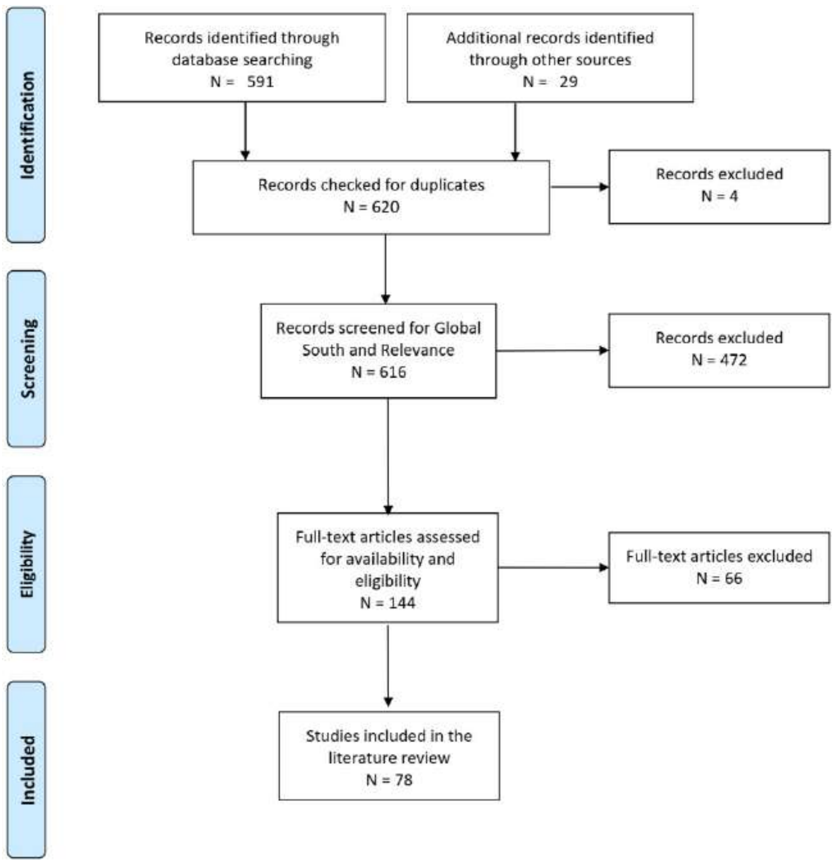


Figure 1. Literature search based on the PRISMA framework.

During the identification, we used the SCOPUS database to search for articles corresponding to our keywords. Based on the research objectives, keywords related to "crowdsourcing," "urban planning," and "Global South" should be identified (Table 1). Drawing from the discussion in Section 2, the main keywords related to "crowdsourcing" are VGI, PPGIS, PGIS, crowdsensing, etc. "Urban planning" was split into two keywords: "urban" (town, city, etc.) and "planning" (management, planning, and policy). The main difficulty was finding keywords related to "Global South," as several studies do not have specific keywords for the GS. Instead, they use the name of the country or city in their abstract, title, or keywords. In order to include as many articles as possible, "Global South" was left out of the keywords and checked during the screening stage.

Table 1. List of possible keywords for each theme.

Crowdsourcing	Urban	Planning
crowd*sourc* participatory sensing crowd*sensing VGI/volunteered geographic information participatory GIS/PGIS/ participatory geographic information system* PPGIS/public participation geographic information system*/ user*generated content	urban residential city/cities/town	planning management policy/policies

Using the keywords displayed in Table 1, we generated the following query:

TITLE-ABS-KEY ((crowd*sourc* OR "participatory sensing" OR crowd*sensing OR vgi OR "volunteered geographic information" OR "participatory gis" OR "participatory geographic information system*" OR "public participation geographic information system*" OR *pgis OR "user*generated content")
AND (urban OR residential OR city OR cities OR town)
AND (planning OR management OR policy OR policies))
AND (LIMIT-TO (SRCTYPE, "j"))
AND (LIMIT-TO (DOCTYPE, "ar"))
AND (LIMIT-TO (LANGUAGE, "English"))

The query above searches for journal articles written in English and corresponding to the keywords described in Table 1. The literature search was first performed in March 2021 and repeated in late December 2021 in order to find the latest articles. In total, 591 articles were obtained from SCOPUS. An additional 29 papers were obtained from other sources (references of selected papers and other reading materials), giving a total of 620 articles.

After removing the duplicates, the remaining articles were screened for relevance and geographic location. The articles corresponding to our research objectives and investigating cities of the GS were retained. A total of 144 articles were obtained at the end of this process.

The available articles from the remaining 144 were downloaded and checked for eligibility based on the following criteria. First, one of the main objectives of this study was to explore the crowdsourcing methods adopted in the GS and their associated challenges. Thus, only studies with a clearly detailed methodology were retained. Second, studies where the data were extracted without the knowledge of the users (social media scraping, non-voluntary crowdsensing, etc.) represented a large portion of the available literature and had to be removed manually. Using the aforementioned criteria, we further screened the database and obtained a final core literature of 78 papers (see Supplementary Materials). The following section provides the descriptive statistics of the reviewed papers.

3.2. Descriptive Statistics

3.2.1. Source Titles and Article Frequency

Table 2 shows that the most represented journals are Remote Sensing (6 articles), Sustainability (5), IEEE Access, Cities, and the International Journal of Geographical Information Science (3). Sustainability, IEEE Access, and PloS One are all Open Access (OA) journals. Furthermore, 25 out of the 78 papers were published in OA journals (about 32%). More OA journals are needed as most researchers in the GS cannot afford journal subscriptions. Open Access journals would be a good way to democratize access to the latest findings and methods in this research area.

Table 2. Top source titles (with at least two articles).

Source Title	Frequency
Remote Sensing	6
Sustainability	5
Cities	3
IEEE Access	3
International Journal of Geographical Information Science	3
GeoJournal	2
Journal of Flood Risk Management	2
Journal of Universal Computer Science	2
PLoS ONE	2

There are also many GIS/engineering journals, which may seem surprising. Given the research topic, one may expect more journals with a planning focus. However, several studies also tried to demonstrate how crowdsourcing could complement other methods, such as remote sensing, to solve the data scarcity problems of the GS [34]. These studies may target non-planning journals such as Remote Sensing Furthermore, several studies tried to optimize the crowdsourcing methods (through new incentive mechanisms, more privacy, better coverage, etc.) using advanced computer and engineering methods [35]. Such studies may target more engineering-oriented journals such as the IEEE series. The presence of GIS journals is mainly due to the fact that several crowdsourcing methods use VGI data (e.g., OSM, web-based PPGIS, etc.). However, all reviewed papers address important urban planning issues and could be of tremendous value for the GS.

All reviewed papers were published in the last 12 years, and the increasing numbers are evidence of a growing interest in the application of crowdsourcing methods in the GS (see Figure 2a).

3.2.2. Large Contribution from China and Researchers Outside the GS

Figure 2b shows that most of the research was conducted by researchers affiliated with Chinese institutes or those outside the GS (United States, United Kingdom, Germany, etc.). In terms of study areas (Figure 2c,d), 38% of the research was conducted in China, followed by 25% in Central and South America (Brazil, Argentina, Guatemala, etc.), 19% in the other parts of Asia (India, Iran, Pakistan, etc.), and 18% in Africa (South Africa, Egypt, Morocco, etc.). There was no contribution from the Pacific Islands.

These numbers show the large domination of China in this research field (both in terms of affiliations and study area). Meanwhile, the other areas of the GS are largely covered by researchers outside the region.

3.2.3. Research Areas

Table 3 shows the main research areas covered in the reviewed papers. We can see that urban morphology and transportation are the most represented areas (16 papers each), followed by environmental monitoring (13 papers). Papers that demonstrate the potential of crowdsourcing as a data collection method, as well as techniques to optimize it and those that assess crowdsourcing tools/methods, represent an important portion of the reviewed papers (9 papers each). Other areas such as urban demographics, disaster detection and management, and smart city management are also covered. The next section provides a detailed description of the main research areas and their key aspects.

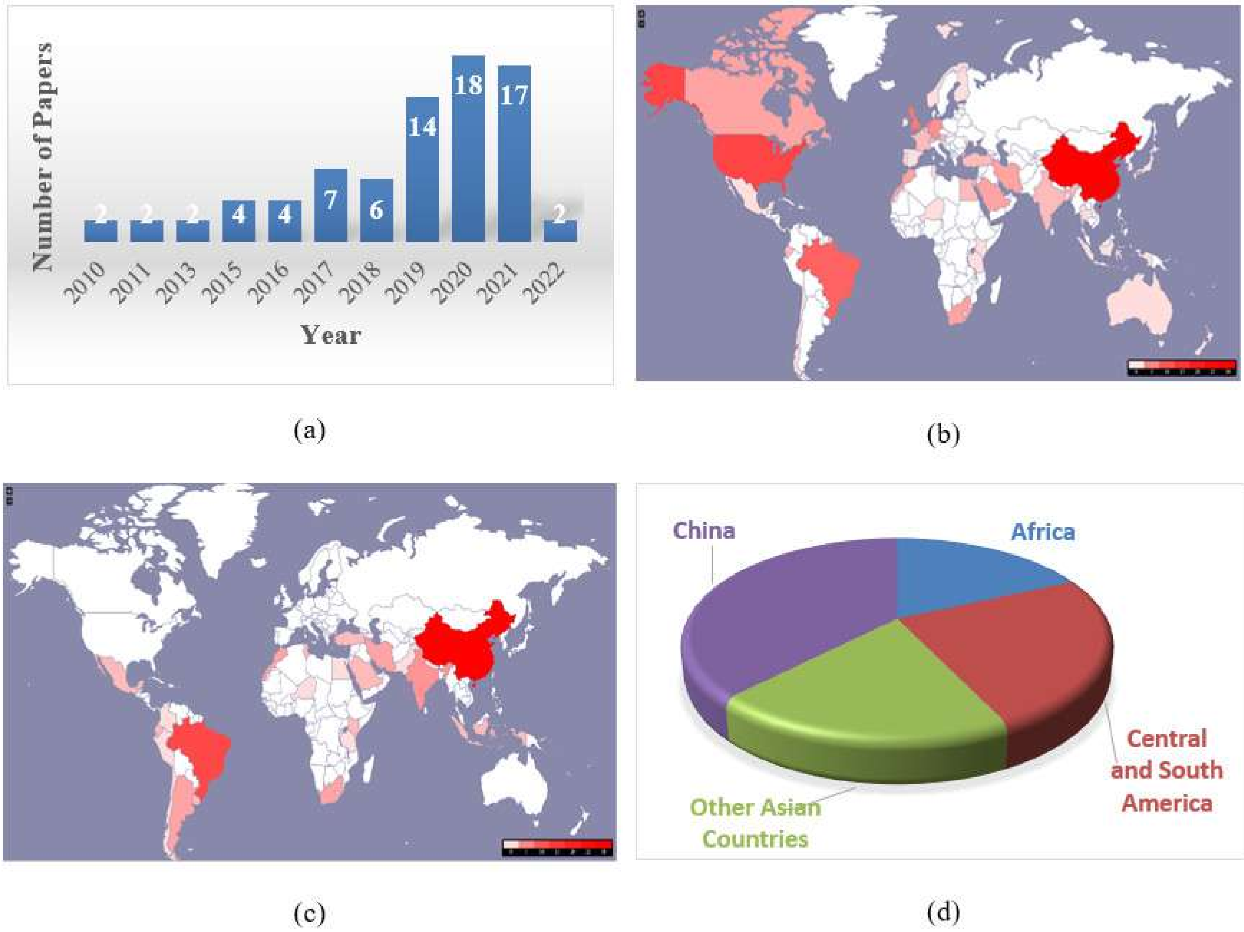


Figure 2. Descriptive statistics. (**a**): Number of articles by year of publication; (**b**): Number of articles by affiliation; (**c**): Number of case studies for each country of the Global South; (**d**): Number of case studies for each area of the Global South and China.

Table 3. Main research areas.

Main Research Areas	Key Aspects	Number of Papers
Urban morphology	Land use, urban landscape, effects of urban forms on physical activities, housing & urban development (neighborhood infrastructure planning, housing schemes, urban development control).	16
Urban transportation	Traffic signal control, public transportation, cycling, traffic flow, intelligent transportation systems (ITS), traffic safety, shared mobility services, parking, multimodal transportation, accessibility, and travel behavior analysis (e.g., route planning, travel pattern analysis, etc.).	16
Environmental monitoring and management	Air quality monitoring (temperature, vehicular emission), sentiment analysis on environmental issues, pollution of coastal zones, waste management, and air quality decision support systems.	13
Urban Data collection and optimization	Introduction of crowdsourcing for urban data collection, optimization of collaborative data collection: distribution, incentive mechanisms, data privacy, data forwarding mechanisms in the context of urban planning.	9
Assessment of Crowdsourcing methods for urban planning	Statistical evaluation of the density, crowd, evolution, and accuracy of crowdsourcing methods (especially collaborative mapping), users' perception of platforms	9
Urban demographics	Urban population estimation (population mapping)	4
Smart city management	Resource management, smart city transformation in the Global South.	4

Table 3. *Cont.*

Main Research Areas	Key Aspects	Number of Papers
Disaster detection and management	Flood detection, mapping and management.	3
Facility location selection	Collaborative selection of facility locations.	1
Public safety	Crime monitoring and management.	1
Urban governance	Citizen participation in urban governance.	1
Urban tourism	Identification of tourism areas of interest.	1

4. Main Research Areas and Keys Aspects

In this section, we use Table 3 to describe the main areas and key aspects covered in the reviewed papers.

4.1. Urban Morphology

These studies use data shared by the public to examine the urban forms, their formation, and evolution, as well as their impact on different aspects of urban life. The main elements of urban forms investigated in the reviewed papers are land use, infrastructures, and housing. The GS experiences fast urbanization which negatively affects the aforementioned elements, and strong measures need to be taken in order to overcome the challenges. In terms of land use, studies in the GS focused on the classification of functional zones so as to determine the main areas where human activities usually occur [36–38]. Such studies are important for the GS as they can help, among others, detect rapid urbanization and can therefore help better manage the existing resources. Crowdsourcing is, in this case, a source of training datasets for the classification algorithms. Regarding infrastructures, they should be a major domain of investigation due to the lack of basic infrastructure in many areas of the GS [39]. Some studies investigated the effects of the road network on cyclist behavior [40]. Studies on urban design focus on the effects of the urban landscape and street configuration on human activities and/or behavior. For example, Mohamed & Stanek [30] examined the effects of street configuration on sexual harassment, while other researchers analyzed the impact of the urban landscape on physical activities [41,42]. Such studies can help guide future urban design so as to build safer, more equitable, and healthier urban environments. Regarding housing, it has been a major cause of concern in the GS, mainly due to the lack of affordable housing and the proliferation of informal settlements. Sub-Saharan Africa has the highest proportion of slums in the world (50.2%), followed by Central and Southern Asia (48.2%) [43]. To tackle these challenges, some studies have involved the public in the mapping of informal settlements in the GS. However, they usually rely on the most basic forms of community mapping with paper drawings and limited sample sizes [44,45]. With the proliferation of smartphones in some parts of the GS, more advanced methods through crowdsourcing could help reach larger samples.

4.2. Urban Transportation

Due to its importance and several implications on different aspects of urban life, transportation is among the most represented areas among the reviewed papers (16 papers). The wide variety of domains covered also justifies the large number of papers in the reviewed literature. As a service designed for the public, transportation is heavily impacted by the way people behave through time and space as well as their response to different transportation-related services. Investigating travelers' behavior could help understand their impact on the urban space (e.g., through their travel patterns) and help draw more data-driven policies to support better transportation planning in the GS. In some cities of the GS, crowdsourcing has been used to examine users' travel behavior through travel patterns [46], route choice [47], travel behavior's impact on congestion [48], etc. Travelers' responses to mobility services as well as strategies to improve them were also investigated. Musakwa and Selala [49] used crowdsourced GPS data to investigate cycling patterns, while other studies developed multimodal or public transportation networks with crowdsourced data [50,51]. Other studies also focus on the traffic signal optimization [52], traffic density estimation [53], etc. Given the large number of social media users among young people, researchers have also looked for ways to involve the youth in transportation planning by crowdsourcing through dedicated social media pages.

4.3. Environmental Monitoring and Management

In an era of sustainable urban planning, research on how public engagement could foster the development of more sustainable cities has become a trend in some cities of the GS. This is also in line with the United Nations' 2030 agenda for sustainable development goals (SDGs) regarding sustainable cities and communities [54], which supports the improvement of urban planning in participatory and inclusive ways. For this reason, researchers have leveraged the power of public engagement through crowdsourcing to monitor the environment and, in some cases, develop decision support systems for both the public and decision-makers. The proliferation of smartphones has made this process easier as smartphones can capture and share data without any technical knowledge from the users. This made possible the collaborative collection of noise data [55], air temperature from smartphone batteries [56,57], the reporting pollution of coastal zones [58], etc.

4.4. Data Collection and Optimization

These studies demonstrate the potential of crowdsourcing as a source of data for the GS as well as ways to optimize the data collection methods. For example, in China, several research efforts have developed new methods to increase the spatio-temporal coverage of voluntary crowdsensing tasks to obtain larger and more representative datasets while minimizing the cost and improving privacy. These methods include protecting participants' privacy, increasing the coverage distribution of sensing tasks through incentive mechanisms [59], and enhancing data forwarding performance through cooperative data forwarding mechanisms [60,61]. Taking into consideration the characteristics of the GS, other studies showed different solutions to involve the public in data gathering and experiment design [62]. Recently, there has been a growing trend on the potential for crowdsourcing as a data collection method for monitoring sustainable development goals (SDGs) in the GS. Pateman et al. [63] provided a review on the use of citizen science for monitoring SDGs in low-and-middle-income countries, while Fraisl et al. [6] introduced a citizen science tool (Picture Pile) for monitoring SDGs.

4.5. Assessment of Crowdsourcing Methods for Urban Planning

Some studies have assessed crowdsourcing methods in the context of urban planning in the GS. Given the novelty of crowdsourcing in the GS, such studies are crucial when assessing its applicability and usefulness for cities in this part of the world. If most studies adopt a more objective approach using statistical evaluations (through the density, accuracy, nature of the crowd, etc.), others opt for a subjective method through users' perceptions (perceived usefulness, perceived ease of use, perceived satisfaction, etc.). The objective assessments mainly focused on collaborative mapping and were conducted in China [64,65], Turkey [66], Kenya [67], as well as cities in Argentina and Uruguay [68], most of them focusing on OSM. Regarding the subjective assessments, Cilliers & Flowerday [69] investigated the subjective factors affecting the intention to use the Interactive Voice Response (IVR) system in South Africa, while Bugs et al. [70] examined the perceived ease of use, perceived usefulness, and satisfaction with a Web-based PPGIS platform for urban planning in Brazil.

4.6. Smart City Management

Smart cities put the public at the center of the planning process. Therefore, participatory approaches such as crowdsourcing play an important role as they allow the public to share their ideas and opinions for more efficient planning practices. However, the GS is behind the rest of the world in terms of smart city management due to a lack of basic infrastructure and a clear understanding of what a smart city should be in local contexts. For this reason, crowdsourcing could start with an exchange on steps towards smart city transformation in the context of the GS. This is the method adopted by Kumar et al. [29], who crowdsourced ideas (idea generation) for smart city transformation in India. Another step would be to consult the public on the efficient management of the existing resources, as demonstrated by other studies in the GS [71].

4.7. Urban Demographics

The rapid population growth in many cities of the GS, especially African cities, raises some challenges which could be mitigated with data-driven methods. Such methods could help monitor the changes in the population, predict future trends and implement proactive policies to face future challenges. However, despite the potential advantages for the GS, urban population estimation has not been widely investigated in the area as all reviewed studies were conducted in China [72–75]. In the aforementioned studies, crowdsourcing (collaborative mapping through OSM) was adopted as supplementary open data so as to improve the accuracy of the mapping algorithms.

4.8. Disaster Detection and Management

If natural disasters are common in all regions of the world, the GS is particularly vulnerable to them due to the lack of resources for disaster detection and management. Crowdsourcing, especially collaborative mapping, has played an important role in helping the GS face these challenges. One of the main examples is the use of OSM for disaster relief during the 2010 earthquake in Haiti. Some studies have shown how public engagement can help improve flood mapping in the GS ([76–78]. Crowdsourced data can supplement other datasets (e.g., wireless sensor networks data) to develop spatial decision support systems (SDSS) for flood management, as demonstrated by Horita et al. [78].

4.9. Other Areas

Some research areas that could have tremendous effects on urban planning have not been widely investigated in the reviewed papers. Although lack of security is often an issue in the GS, only one study has addressed it among the reviewed papers [79]). This is also the case for urban tourism, urban governance, and facility location selection.

Several studies that were excluded also discussed different aspects of urban planning using social media data. As we explained in Section 2, social media without explicit consent from the crowd are excluded from the reviewed literature. The topics discussed in those studies include urban health (e.g., COVID mapping in urban areas), urban tourism, disaster detection and management (earthquake, flood detection, etc.).

5. Crowdsourcing Methods

This section provides an overview of the crowdsourcing methods applied in the GS. More specifically, for each method, we discuss the basic principles, its potential value for urban planning, especially in the context of the GS, the type of data obtained and, finally, its main areas of application. The methods identified in this study can be regrouped into three categories: collaborative websites, voluntary crowdsensing, and dedicated social media campaigns.

5.1. Collaborative Websites

Collaborative websites are web-based platforms that allow participants to share local knowledge, maps, geo-tagged pictures, etc., within a specific framework. Globally, they have the benefit of providing a participatory planning process that allows the end users to comment on planning projects [70], map the areas they are most familiar with (e.g., OSM), report violations and crimes [80], suggest innovative ideas for smart city planning [23], etc. Collaborative websites are generally in line with the concept of collective intelligence [81], as individual knowledge is openly available to other participants who can access, edit, discuss and improve them. As a result, they yield better outcomes than knowledge from single individuals. In the GS, government agencies usually lack the equipment and workforce to adequately handle the many problems they face on a daily basis [58,80]. Collaborative websites can provide a low-cost and effective solution to these problems while raising awareness among the people.

For more clarity, we distinguish between mapping as a separate endeavor and providing user-generated content to support a specific area of planning (crime reporting, environmental monitoring, etc.) using web-based PGIS/PPGIS, for example. This is because some maps are collaboratively built for general purposes (e.g., OSM). The resulting map could then be applied to different areas, including urban planning. The next subsection discusses collaborative mapping as a separate endeavor.

5.1.1. Collaborative Mapping

Collaborative mapping is a collective effort in which volunteers with various levels of expertise and motivations participate in the creation, edition, and dissemination of digital maps [27,82]. The mapping process relies on the collection (through sensors), assemblage, and annotation of geographic information using web mapping tools (e.g., OSM website). Since these web mapping tools are easy to use, even non-experts can take part in collaborative mapping, which helps reach a potentially larger crowd.

These mapping endeavors have several potential advantages for the GS, including accessibility and accuracy. Accessibility refers to the possibility for any user to freely use the maps. OSM provides this possibility as long as the user acknowledges their use of the service [82]. Furthermore, many areas of the GS still use outdated maps due to the costliness of professional mapping services [67]. Collaborative platforms offer the same level of accuracy as commercial mapping services [83]. These factors make collaborative mapping a potential source of reliable and up-to-date urban data with minimal cost for areas with limited resources, such as the GS.

The main data used from collaborative mapping in the GS consists of road networks (road types, stations, etc.) [47,84,85]. Datasets related to building footprint have also been used [86]. Collaborative mapping has been used to investigate urban morphology [40], land use [84,87], transportation planning [47,50], urban population estimation [74], etc. Recently, there has been an increasing trend in the contributions of corporate editors to OSM, especially in Southeast Asia [88]. Finally, it is worth noting that some collaborative mapping platforms also apply to specific areas such as disaster relief, election monitoring, etc. One example of such platforms in the GS is the use of the Ushahidi platform for flood mapping in Brazil [77].

5.1.2. Web-Based PPGIS

A web-based PPGIS framework consists of four main concepts: GIS, public participation, web development, and the domain of application (e.g., crime monitoring, environmental monitoring, flood mapping, etc.). As such, it is a multidisciplinary area that allows participants to share local knowledge through online GIS platforms. Participants can use the platform to post or comment on different urban problems, including infrastructure damage, crime, natural disasters, etc. The importance of the aforementioned problems and the risk they could represent in a community are among the main factors that explain the need for the public to actively participate in these PPGIS projects. Furthermore, artificial sensors (i.e., cameras), which are often used to monitor crime and other types of violations do not have the intelligence to provide an in-depth and real-time interpretation of the events. In this case, the public could provide a better response to the issue (e.g., crime reporting, helping the victim, etc.) than an artificial sensor. The posts can be in the form of geotagged text, audio, video, or a combination. The fact that other participants can also comment on a post helps ensure the reliability of the information provided.

Web-based PPGIS has been adopted in the GS for environmental monitoring [58], housing [89], crime management and monitoring [79], flood risk management [78], etc.

There are also bottom-up decision support systems that allow the public to get involved throughout the decision-making process. In Iran, the web-based Spatial Decision Support System (WebGIS-SDSS) allows citizens to access, discuss, review and submit their opinions about urban development applications based on multi-criteria decision-making [90]. The authorities then aggregate the opinions to make their final decision (accept or reject the application).

5.1.3. Idea Generation/Idea Contest

Some collaborative websites provide a platform for innovative ideas to support modern and sustainable city planning. They allow non-experts to actively participate in the planning process by identifying their needs, submitting innovative ideas, commenting and/or rating other users' suggestions [23], or helping choose the location of future facilities. Unlike collaborative mapping and web-based PPGIS, these websites do not necessarily require the use of geo-location services, and users can directly participate without any prior GIS knowledge. Despite its potential advantages for citizen empowerment, this type of collaborative website was rare among the reviewed papers. Examples in the GS include idea generation for smart city transformation in India [29] and facility location selection [91].

5.2. Voluntary Crowdsensing

The latest smartphones are equipped with a variety of sensors, including cameras, accelerometers, microphones, a global positioning system (GPS), air quality sensors, etc. Furthermore, there have been tremendous improvements in the memory size and computational capabilities of these mobile devices in the last few years. These factors, coupled with the increasing accessibility of smartphones, have turned mobile phone owners from simple users to contributors of rich sensing data. In addition to smartphones, these sensors can also be installed in other devices (laptops, tablets, etc.) or locations such as cars and help gather data for traffic or environmental monitoring from a potentially large group of participants. Crowdsensing uses the power of the crowd and the ubiquity of sensing technologies to collect data for various urban planning activities, including traffic management, environmental monitoring, etc.

Crowdsensing has advantages in investigating modes with low share, such as cycling. Given the small number of cyclists in many cities (between 1 and 2% for work trips), it may be difficult to find a representative sample for analyzing cycling patterns using traditional methods such as cycle count [92]. In this case, crowdsensing could help cover a larger and more diverse sample (e.g., Strava in Johannesburg, South Africa [49]). Crowdsensing is also beneficial in traffic control and management. It could be a time and cost-efficient alternative to roadside cameras and loop detectors to detect traffic congestion. Most developing countries cannot afford these cameras or other roadside sensors and could highly benefit from these methods [52]. In Africa, recent studies have used GPS devices to address the lack of reliable data [93,94]. However, these methods are expensive and suffer from a limited sample size [93,94]. Crowdsensing could provide a low-cost solution to these problems [95].

Data from crowdsensing usually consists of GPS tracks, temperature, noise level, particulate matter ($PM_{2.5}$ and PM_{10}), geotagged pictures/videos/audio/comments shared by participants, etc. Crowdsensing can also provide large urban datasets for planners. Examples of such datasets in the GS include datasets for environmental sensing [55,57], GPS data for cycling patterns analysis [49]), and smart city management [96], Waze data for urban mobility [48], etc.

As specified in Section 2, we only investigated voluntary participation in crowdsensing throughout our analysis.

5.3. Dedicated Social Media Campaigns

Social media campaigns (SMC) are not to be confused with social media scraping (through Twitter or Weibo). By SMC, we refer to specific social media pages, groups, etc.

which are launched with a clear goal (e.g., addressing an urban planning issue), the task is clearly defined and outsourced *to* the crowd (submission of proposals, comments, complaints, vote, etc.), and the participation is always voluntary. They are, to some extent, similar to collaborative websites. However, unlike many collaborative websites which require technical skills and/or financial resources (that are not always available in the GS), SMC leverage existing social media platforms and are easy to set up and manage. A concrete example would be the use of a dedicated Facebook group to discuss public safety in an urban area. The crowdsourcer can launch a page that invites all inhabitants of a specific area to report issues, comment, and suggest solutions. The information provided could be valuable to policymakers and the public without any financial burden.

Examples of applications in the GS include Facebook pages for public participation in transportation planning [97] and e-government [98], sexual harassment reporting on Facebook and Twitter [30], etc.

6. Discussion

6.1. Challenges

Although some challenges, such as sample representativeness, privacy, access, and data processing, are applicable to all crowdsourcing projects [99], some issues are specific to or more severe in the GS. The challenges discussed here are drawn from the characteristics of the GS and corroborated with the 78 reviewed papers.

6.1.1. The Digital Divide

Despite the recent improvements, the digital divide is still present in many parts of the world [63]. The GS is also characterized by a technological gap with respect to the rest of the world. For example, in 2022, Internet penetration rates are lower in Africa (43.2%), The Caribbean/South America (80.5%), the Middle East (77.1%) and Asia (67.0%) [100]. In comparison, North America has a 93.4% penetration rate, while 89.2% of Europeans have access to the Internet. Furthermore, access to mobile Internet is also lower in low and middle-income countries [101]. Given the importance of the Internet in crowdsourcing, the GS suffers a severe disadvantage compared with the rest of the world. In addition to access, the ability to use the technology is also an indicator of the digital divide. A lack of literacy and digital skills is a barrier to mobile Internet use in low-and-middle-income countries [101]. As a result, many people lack the knowledge or means to effectively handle the technology needed to participate in collaborative mapping [27] or web-based planning decision support systems. This is evidenced by Young et al. [102], whose crowdsourcing effort across Africa was hindered by slow and expensive Internet connections, regular disruptions due to power outages, and participants' limited digital skills. Zia et al. [66] also found a strong correlation between literacy level and the density of OSM in Turkey. Thus, the digital divide could lead to limited mapping coverage and/or reliance on armchair mappers. Despite their efforts, armchair mappers lack local knowledge, which can have significant effects on the accuracy of the produced maps. De Leeuw et al. [67] found that participants with local knowledge (including laypeople) achieved significantly higher accuracies than those without local knowledge, including professional mappers.

6.1.2. Academic Challenges and Digital Colonialism

Figure 2b shows that, besides China, most of the studies were conducted by researchers outside the GS. The possible reasons for the limited numbers of researchers from the GS (besides China) are a lack of access to the technologies (see Section on digital divide), limited resources for undertaking data collection campaigns, and a lack of trained experts able to process the data. For these reasons, countries of the GS are at the mercy of NGOs, funding agencies, and institutes of the Global North whose interests may not be aligned with the challenges faced by cities of the GS. This predominance of foreign entities could be an opportunity if they fully involved their local counterparts in the process. However, since foreign institutes and NGOs also provide the funding, their collaboration with local scholars usually turns into a top-down relationship where local researchers are merely used as "glorified data collectors" [103]. This hierarchical relationship affects the way research is conducted in the GS and could leave out important issues that affect the local people. For example, a few studies among the reviewed literature address issues related to public safety (e.g., crime mapping), illegal dumping, gender-related issues, or lack of basic facilities and services in many GS cities (e.g., good roads, poor public transportation systems, etc.). Beyond the academic area, this also raises several questions about the possible "exploitation" of the GS citizens whose efforts to contribute data to crowdsourcing campaigns may only serve the interests of foreign (usually Global North) organizations. Do the data contributed result in solutions that help the participants? Do the participants have access to the data they contributed? Do these external scholars and NGOs give an accurate and unbiased portrayal of the Global South? All these issues have raised concerns over new forms of "digital and data colonialism" [102,104]. Moreover, the growing influence of corporate editors [88] could be problematic, especially in the vulnerable areas of the GS, if the generated maps serve corporate interests rather than the local people. Digital colonialism poses challenges to citizen empowerment, data ownership, and academic excellence. If these challenges are not addressed, they could contribute to enforcing the same North-South inequalities that democratic processes such as crowdsourcing were supposed to mitigate.

It is important to clarify that we did not see any direct relationship between a particular crowdsourced method and the former colonial ruler. This is due to the fact that the crowdsourcing projects are usually launched by different Western organizations regardless of who the former colonial power was. For example, a project in Niger (a former French colony) was initiated by a German organization [105]). Another project in Mexico (a former Spanish colony) was initiated by a Belgian organization [62]. All these projects have certain characteristics in common.

- Besides China, most projects were initiated by foreign, western universities, an indication of dependence on western countries for crowdsourcing.
- Such dependence has implications in terms of data ownership, research design, and administration (as we explained in Section 6), which lead to the phenomenon of digital colonialism.

Thus, digital colonialism is not necessarily associated with a particular former colonial ruler; it is due to the presence and practices of western organizations whose control over the data may not serve the interests of the local communities.

It is, however, important to point out that digital colonialism's impact on the global south varies depending on the region. As shown in Figure 2b, China is less dependent on foreign researchers.

6.1.3. Socio-Economic and Cultural Challenges

The issues raised here are related to income, age, and gender.

Regarding age and income, about 25% of adults in low-and-middle-income countries are unaware of mobile Internet, while more than half of the people do not meet the UN's mobile Internet affordability target [101]. Lack of awareness and/or affordability of mobile Internet is an obstacle to people's participation in mobile crowdsourcing and a source of bias. The potential cultural issues are mostly related to gender. Gender issues in the GS include the assumption in many cultures that women lack the ability to provide useful information in mapping projects, for example [106]. Furthermore, women from low-and-middle-income countries are 20% less likely than men to use mobile Internet. Since one of the main objectives of crowdsourcing is to democratize the planning process and empower the public, all stakeholders need to actively and freely participate regardless of gender. In addition, crowdsourcing could help address many of the problems (primarily) faced by women, such as sexual harassment. One example of such a case in the GS is HarassMap, a platform for reporting sexual harassment in Egypt. Although such platforms could help raise awareness and encourage the authorities to address this issue, the authors also pointed out the lack of female participation as a major challenge [30]. Another cultural challenge in the GS is the expectation of financial reward in exchange for participation, even when the project has clear value for the community [102]. This could seriously limit the number of participants as many citizen engagement projects do not offer any financial reward.

6.1.4. Administrative Challenges

Public participation should integrate the input of all stakeholders into the planning process. In this regard, citizens should not be simple providers of census data or travel diaries, nor should their role be limited to a simple consultation prior to decision making. Instead, they should be involved throughout the decision-making process. However, the GS is mostly characterized by top-down governance systems, which give little to no room for full citizen participation [90]. This is also evidenced in the literature corpus analyzed in this study, where a few studies provide platforms for citizen participation in problem-solving or decision-making.

The existence of top-down governance systems also does not allow to break away from traditional authoritative data collection methods. Many projects often seek approval from local authorities before implementation [102].

6.2. Suggestions for Future Implementations

6.2.1. Data Ownership and Benefits for the Public: A Solution to Digital Colonialism and Low Participants' Motivation

It is important for the public to know that their efforts to collect and/or share data/information will be accessible to them, or they will result in solutions, technologies, or policies that will help them, not exploit them. A failure to meet the aforementioned requirements could further reinforce digital colonialism and affect the public's motivation to participate in any crowdsourcing initiative [102]. In the GS, some scholars address this issue by looking for ways to help the public directly appropriate their data and/or participate in the interpretation of the results. For example, in Niamey (Niger, Africa), participants and their families were also involved in the analysis and interpretation of the crowdsourced photos [105]. In China, Li et al. [55] stated that data shared by the public through their collaborative environmental sensing network (CESN) would be publicly available, which could increase the number of participants. Unlike Strava Metro, which is a commercial platform, OSM data is available to the public. Thus, participants in OSM projects can directly access the fruit of their labor and use it for future endeavors. This makes OSM more suitable for the GS than other commercial services. Experts should also use the data contributed by the public to develop solutions that will benefit them. In Kenya, Williams et al. [107] used the data collected through the digital Matatu project to help local experts build a mobile crowdsourcing application (Ma3Route), which shares real-time traffic data with users [108]. In Morocco, El Alaoui El Abdallaoui et al. [109] designed an air quality decision support system that uses collaborative environmental sensing data to recommend the least polluted route and display information pertaining to public health. Finally, it is important to get participants more involved in the design of the methodologies. This could help design methods that are more suitable for the participants and ensure that the projects' goals are in line with community priorities [63]. An example is provided in Mexico, where participants and experts codesigned the crowdsourcing experiment for SenseCityVity [62].

6.2.2. For Governments and Research Institutes

In addition to the public, digital colonialism also affects local scholars. Crowdsourcing could be an opportunity for strong collaboration between North-South researchers if challenges related to digital and data colonialism are addressed. To do so, a new framework that integrates equal inputs from both sides is needed. More concretely, local researchers should not be "exploited" for data collection. Instead, they should fully participate in the definition of the objectives and methodology so that the research is in line with the needs of the GS and does not repeat the same North-South inequalities that characterize digital and data colonialism.

This review has shown that there is an urgent need for more research in many areas. The Caribbean, Central America, the Pacific Islands, the Middle East, and Sub-Saharan Africa are barely covered, with the research largely condensed in China and Brazil. This opens an opportunity for more research and perhaps more insight into the urban dynamics of this area of the world that remains unknown. Furthermore, as discussed in Section 4, several topics have not been widely investigated despite their importance for the GS and the potential benefits of public engagement in these areas. This shows the tremendous potential that is yet to be fulfilled in the application of crowdsourcing for urban studies in the GS.

For governments, it is important to include public participation in all aspects of their programs to raise awareness and encourage citizen engagement. A good example in the GS is Brazil, which encourages public engagement via legislation [70]. This seems to have a major effect as a large part of the platforms for public engagement were found in Brazil [70,71,77,78,110].

Finally, authorities, as well as researchers, should adopt more affordable methods such as open-source software (which allows access and replication of the methods) [111], as well as accessible and easy-to-use platforms such as Ushahidi (https://www.ushahidi.com/about/pricing, accessed on 9 May 2022). Since most crowdsourcing projects are not sustained due to a lack of resources, these solutions could be useful to the GS.

6.2.3. Solutions to Socio-Economic and Cultural Challenges

Sample representativeness is a challenge in all crowdsourced projects. The problem is bigger in the GS, where some segments of the population are excluded due to local customs (see Section 6.1.3). Including all members of the society helps mitigate sample bias and provide a more accurate analysis of the problem under study, which in turn helps implement more suitable policies. A change of mentality (especially regarding women) is required. Another way to include women is to adapt the public engagement process to their temporal and spatial constraints, for example, by allowing them to participate with their children [63].

Other solutions to socio-economic and cultural barriers include using emojis, pictures, checkmarks, and voice messages to reach the illiterate [63,69,111]. An interesting example in the GS would be the adoption of the interactive voice response (IVR) system in South Africa [69]. The IVR is a crowdsourcing system that allows users to report public safety concerns by directly recording and sharing a voice message without having the write or deal with a web interface. Furthermore, the IVR system can be available through a toll-free number (accessible to the poor) and be adapted to local languages.

Finally, the use of mixed methods is also important. This includes combining crowdsourcing with traditional methods to obtain more representative samples. For example, in areas with low literacy levels, the project could involve calling, chatting on social media platforms (WhatsApp, for example), and having personal meetings with the participants in order to explain in detail the basics of the participation process [102] or combining the crowdsourced data with census data.

7. Conclusions

Although crowdsourcing has several advantages, it is important for planners to have a clear understanding of the target population [15]. This could help anticipate some of the challenges that could affect the quality of the crowdsourced data. Given the unique characteristics of the GS, it was thus crucial to conduct a review of the crowdsourcing methods adopted in this region of the world and highlight their potential benefits and challenges so as to provide some suggestions for future research.

To achieve this goal, we reviewed 78 English-written journal articles focusing on voluntary participation in crowdsourcing in the GS. The reviewed articles were mainly contributed by researchers affiliated with Chinese institutes or outside the GS. Among the crowdsourcing methods, collaborative mapping (through OSM) was most widely adopted, while the studies covered a variety of areas, including urban transportation, event detection and crisis management, urban tourism, urban health, environmental monitoring, gender, etc. Based on the descriptive statistics of the reviewed papers and the characteristics of the GS, we discussed the potential administrative, academic, technological, socio-economic, and cultural challenges that could affect the successful adoption of crowdsourcing in the GS. Solutions to these challenges were provided as suggestions for future implementations and included new collaboration frameworks with foreign experts so as to avoid digital colonialism, the inclusion of all segments of the population (especially women), the use of more accessible platforms to foster public participation in urban planning (e.g., Ushahidi), the development of methods that are more in line with the needs and the characteristics of the GS, etc.

Overall, this study has demonstrated that even though crowdsourcing has been heralded as a means for less developed countries to gather large urban data at a minimal cost and foster citizen empowerment and awareness through public platforms, several challenges still need to be addressed. The needed datasets and/or platforms are VGI (e.g., OSM) datasets to complement remote sensing datasets for investigating the challenges of the GS (e.g., informal settlements, lack of data for disaster response, crimes, lack of clean water, etc.), citizen sensing data for a better understanding of mobility patterns (e.g., GPS) and environmental monitoring (noise, temperature, etc.), recommendations/solutions for better planning practices, etc.

This study is, however, not without limitations. The inclusion of China in the literature could be misleading as a large portion of the reviewed papers is from this country. This might give the impression that the GS, in general, has produced a large part of the studies on this topic. The main reason for including China was the fact that, despite its rapid growth in the last four decades, the country still displays some characteristics of the GS [112]. Thus, its inclusion could help compare it with the rest of the GS and stress the progress that is to be made in order to perhaps reach the same standards in terms of innovation and academic achievements. Furthermore, although a careful and well-justified literature search was adopted, the small number of studies from French-speaking African countries and Latin America could be due to the fact that only articles written in English were included in the review. Given the large number of French-speaking countries in Africa and Spanish-speaking countries in Latin America, some valuable contributions could have been left out. English is, however, the language adopted in most studies and literature reviews [113], and we believe that reviewed articles offer a clear overview of the crowdsourcing methods adopted for urban planning in the GS as well as the associated challenges and our suggestions are expected to encourage more research in this area.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su141811461/s1 File S1: List of Reviewed Papers.

Author Contributions: Conceptualization, E.B.D., J.C., S.C.K.T. and R.A.; methodology, E.B.D.; software, E.B.D.; validation, E.B.D., J.C., S.C.K.T. and R.A.; formal analysis, E.B.D.; data curation, E.B.D.; writing—original draft preparation, E.B.D.; writing—review and editing, E.B.D., J.C., S.C.K.T. and R.A.; supervision, J.C.; project administration, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available in File S1.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Howe, J. The Rise of Crowdsourcing. *Wired Mag.* **2006**, *41*, 1–4. Available online: https://www.wired.com/2006/06/crowds/ (accessed on 9 May 2022).
2. Toyota Global. *Data: Changes in Toyota Trademarks and Emblems*; Toyota Global: Toyota, Japan, 2012.
3. Davidoff, P. Advocacy and Pluralism in Planning. *J. Am. Inst. Plann.* **1965**, *31*, 331–338. [CrossRef]
4. UN ESCAP. *Asia and the Pacifc SDG Progress Report 2020*; United Nations: Bangkok, Thailand, 2020.
5. Fraisl, D.; Campbell, J.; See, L.; Wehn, U.; Wardlaw, J.; Gold, M.; Moorthy, I.; Arias, R.; Piera, J.; Oliver, J.L.; et al. Mapping Citizen Science Contributions to the UN Sustainable Development Goals. *Sustain. Sci.* **2020**, *15*, 1735–1751. [CrossRef]
6. Fraisl, D.; See, L.; Sturn, T.; MacFeely, S.; Bowser, A.; Campbell, J.; Moorthy, I.; Danylo, O.; McCallum, I.; Fritz, S. Demonstrating the Potential of Picture Pile as a Citizen Science Tool for SDG Monitoring. *Environ. Sci. Policy* **2022**, *128*, 81–93. [CrossRef]
7. Njoh, A.J. The Experience and Legacy of French Colonial Urban Planning in Sub-Saharan Africa. *Plan. Perspect.* **2004**, *19*, 435–454. [CrossRef]
8. United Nations. *How Building Codes and Regulations Can Be Adapted to Meet the Basic Needs of the Poor: Report of the UN Seminar of Experts on Building Codes and Regulations in Developing Countries, Tällberg and Stockholm, March, 1980*; Swedish Council for Building Research: Stockholm, Sweden, 1980; ISBN 91-540-3251-2.
9. Chenal, J. Les Villes Africaines en Quête de Nouveaux Modèles Urbanistiques. Available online: https://metropolitiques.eu/Les-villes-africaines-en-quete-de.html (accessed on 10 January 2021).
10. Insua, R.D.; Kersten, E.G.; Rios, J.; Grima, C. Towards Decision Support for Participatory Democracy. *ISeB* **2008**, *6*, 161–191. [CrossRef]
11. Bai, S.; Jiao, J. From Shared Micro-Mobility to Shared Responsibility: Using Crowdsourcing to Understand Dockless Vehicle Violations in Austin, Texas. *J. Urban Aff.* **2020**, *42*, 1–13. [CrossRef]
12. Thiagarajan, A.; Biagioni, J.; Gerlich, T.; Eriksson, J. Cooperative Transit Tracking Using Smart-Phones. In Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems, Zurich, Switzerland, 3–5 November 2010; pp. 85–98.
13. De Filippi, F.; Coscia, C.; Boella, G.; Antonini, A.; Calafiore, A.; Cantini, A.; Guido, R.; Salaroglio, C.; Sanasi, L.; Schifanella, C. MiraMap: A We-Government Tool for Smart Peripheries in Smart Cities. *IEEE Access* **2016**, *4*, 3824–3843. [CrossRef]
14. Hu, K.; Sivaraman, V.; Luxan, B.G.; Rahman, A. Design and Evaluation of a Metropolitan Air Pollution Sensing System. *IEEE Sens. J.* **2016**, *16*, 1448–1459. [CrossRef]
15. Griffin, G.P.; Jiao, J. The Geography and Equity of Crowdsourced Public Participation for Active Transportation Planning. *Transp. Res. Rec.* **2019**, *2673*, 1–9. [CrossRef]
16. Kemajou, A.; Konou, A.A.; Jaligot, R.; Chenal, J. Analyzing Four Decades of Literature on Urban Planning Studies in Africa (1980–2020). *Afr. Geogr. Rev.* **2021**, *40*, 425–443. [CrossRef]
17. Kong, X.; Liu, X.; Jedari, B.; Li, M.; Wan, L.; Xia, F. Mobile Crowdsourcing in Smart Cities: Technologies, Applications, and Future Challenges. *IEEE Internet Things J.* **2019**, *6*, 8095–8113. [CrossRef]
18. Kanhere, S.S. Participatory Sensing: Crowdsourcing Data from Mobile Smartphones in Urban Spaces. In Proceedings of the IEEE International Conference on Mobile Data Management, Lulea, Sweden, 6–9 June 2011; Volume 2, pp. 3–6.
19. Niu, H.; Silva, E.A. Crowdsourced Data Mining for Urban Activity: Review of Data Sources, Applications, and Methods. *J. Urban Plan. Dev.* **2020**, *146*, 04020007. [CrossRef]
20. Certomà, C.; Corsini, F.; Rizzi, F. Crowdsourcing Urban Sustainability. Data, People and Technologies in Participatory Governance. *Futures* **2015**, *74*, 93–106. [CrossRef]
21. Criscuolo, L.; Carara, P.; Bordogna, G.; Pepe, M.; Zucca, F.; Seppi, R.; Ostermann, F.; Rampini, A. *Handing Quality in Crowdsourced Geographic Information.*; Ubiquity Press Ltd.: London, UK, 2016.
22. Wang, X.; Zheng, X.; Zhang, Q.; Wang, T.; Shen, D. Crowdsourcing in ITS: The State of the Work and the Networking. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 1596–1605. [CrossRef]
23. Schuurman, D.; Baccarne, B.; De Marez, L.; Mechant, P. Smart Ideas for Smart Cities: Investigating Crowdsourcing for Generating and Selecting Ideas for ICT Innovation in a City Context. *J. Theor. Appl. Electron. Commer. Res.* **2012**, *7*, 49–62. [CrossRef]
24. Estellés-Arolas, E.; González-Ladrón-De-Guevara, F. Towards an Integrated Crowdsourcing Definition. *J. Inf. Sci.* **2012**, *38*, 189–200. [CrossRef]
25. Brabham, D.C. Crowdsourcing the Public Participation Process for Planning Projects. *Plan. Theory* **2009**, *8*, 242–262. [CrossRef]
26. Kumar, H.; Singh, M.K.; Gupta, M.P. Smart Mobility: Crowdsourcing Solutions for Smart Transport System in Smart Cities Context. In Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance, Galway, Ireland, 4–6 April 2018; pp. 482–488.
27. Heipke, C. Crowdsourcing Geospatial Data. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 550–557. [CrossRef]
28. Nakatsu, R.T.; Grossman, E.B.; Iacovou, C.L. A Taxonomy of Crowdsourcing Based on Task Complexity. *J. Inf. Sci.* **2014**, *40*, 823–834. [CrossRef]
29. Kumar, H.; Singh, M.K.; Gupta, M.P.; Madaan, J. Moving towards Smart Cities: Solutions That Lead to the Smart City Transformation Framework. *Technol. Forecast. Soc. Chang.* **2020**, *153*, 119281. [CrossRef]
30. Mohamed, A.A.; Stanek, D. The Influence of Street Network Configuration on Sexual Harassment Pattern in Cairo. *Cities* **2020**, *98*, 102583. [CrossRef]

31. Guo, B.; Yu, Z.; Zhou, X.; Zhang, D. From Participatory Sensing to Mobile Crowd Sensing. In Proceedings of the 2014 IEEE International Conference on Pervasive Computing and Communication Workshops, Percom Workshops, Budapest, Hungary, 24–28 March 2014; pp. 593–598.
32. Nummi, P. Crowdsourcing Local Knowledge with PPGIS and Social Media for Urban Planning to Reveal Intangible Cultural Heritage. *Urban Plan.* **2018**, *3*, 100–115. [CrossRef]
33. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* **2009**, *6*, e1000097. [CrossRef] [PubMed]
34. Forget, Y.; Shimoni, M.; Gilbert, M.; Linard, C. Mapping 20 Years of Urban Expansion in 45 Urban Areas of Sub-Saharan Africa. *Remote Sens.* **2021**, *13*, 525. [CrossRef]
35. Yu, Z.; Han, L.; An, Q.; Chen, H.; Yin, H.; Yu, Z. Co-Tracking: Target Tracking via Collaborative Sensing of Stationary Cameras and Mobile Phones. *IEEE Access* **2020**, *8*, 92591–92602. [CrossRef]
36. Liu, X.; He, J.; Yao, Y.; Zhang, J.; Liang, H.; Wang, H.; Hong, Y. Classifying Urban Land Use by Integrating Remote Sensing and Social Media Data. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 1675–1696. [CrossRef]
37. Ye, Y.; An, Y.; Chen, B.; Wang, J.J.; Zhong, Y. Land Use Classification from Social Media Data and Satellite Imagery. *J. Supercomput.* **2020**, *76*, 777–792. [CrossRef]
38. Xing, H.; Meng, Y. Integrating Landscape Metrics and Socioeconomic Features for Urban Functional Region Classification. *Comput. Environ. Urban Syst.* **2018**, *72*, 134–145. [CrossRef]
39. Dempsey, N.; Brown, C.; Raman, S.; Porta, S.; Jenks, M.; Jones, C.; Bramley, G. Elements of Urban Form. In *Dimensions of the Sustainable City*; Jenks, M., Jones, C., Eds.; 2008; pp. 21–51, ISBN 9781402086472.
40. Orellana, D.; Guerrero, M.L. Exploring the Influence of Road Network Structure on the Spatial Behaviour of Cyclists Using Crowdsourced Data. *Environ. Plan. B Urban Anal. City Sci.* **2019**, *46*, 1314–1330. [CrossRef]
41. Ma, M.; Ding, L.; Kou, H.; Tan, S.; Long, H. Effects and Environmental Features of Mountainous Urban Greenways (MUGs) on Physical Activity. *Int. J. Environ. Res. Public. Health* **2021**, *18*, 8696. [CrossRef]
42. Liu, K.; Siu, K.W.M.; Gong, X.Y.; Gao, Y.; Lu, D. Where Do Networks Really Work? The Effects of the Shenzhen Greenway Network on Supporting Physical Activities. *Landsc. Urban Plan.* **2016**, *152*, 49–58. [CrossRef]
43. Statista Share of Urban Population Living in Slums in 2020, by Region. Available online: https://www.statista.com/statistics/684694/percentage-of-world-urban-population-in-slums-by-region/ (accessed on 1 August 2022).
44. Panek, J.; Sobotova, L. Community Mapping in Urban Informal Settlements: Examples from Nairobi, Kenya. *Electron. J. Inf. Syst. Dev. Ctries.* **2015**, *68*, 1–13. [CrossRef]
45. Vergara-Perucich, F.; Arias-Loyola, M. Community Mapping with a Public Participation Geographic Information System in Informal Settlements. *Geogr. Res.* **2021**, *59*, 268–284. [CrossRef]
46. Pedreira Junior, J.U.; Assirati, L.; Pitombo, C.S. Improving Travel Pattern Analysis with Urban Morphology Features: A Panel Data Study Case in a Brazilian University Campus. *Case Stud. Transp. Policy* **2021**, *9*, 1715–1726. [CrossRef]
47. Wu, T.; Zeng, Z.; Qin, J.; Xiang, L.; Wan, Y. An Improved Hmm-Based Approach for Planning Individual Routes Using Crowd Sourcing Spatiotemporal Data. *Sensors* **2020**, *20*, 6938. [CrossRef]
48. Calatayud, A.; Sánchez González, S.; Marquez, J.M. Using Big Data to Estimate the Impact of Cruise Activity on Congestion in Port Cities. *Marit. Econ. Logist.* **2022**, *24*, 566–583. [CrossRef]
49. Musakwa, W.; Selala, K.M. Mapping Cycling Patterns and Trends Using Strava Metro Data in the City of Johannesburg, South Africa. *Data Brief* **2016**, *9*, 898–905. [CrossRef]
50. Frez, J.; Baloian, N.; Pino, J.A.; Zurita, G.; Basso, F. Planning of Urban Public Transportation Networks in a Smart City. *J. Univers. Comput. Sci.* **2019**, *25*, 946–966.
51. Smarzaro, R.; Davis, C.A.; Quintanilha, J.A. Creation of a Multimodal Urban Transportation Network through Spatial Data Integration from Authoritative and Crowdsourced Data. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 470. [CrossRef]
52. Dixit, V.; Nair, D.J.; Chand, S.; Levin, M.W. A Simple Crowdsourced Delay-Based Traffic Signal Control. *PLoS ONE* **2020**, *15*, e0230598. [CrossRef] [PubMed]
53. Huang, Y.; Tian, Y.; Liu, Z.; Jin, X.; Liu, Y.; Zhao, S.; Tian, D. A Traffic Density Estimation Model Based on Crowdsourcing Privacy Protection. *ACM Trans. Intell. Syst. Technol.* **2020**, *11*, 1–8. [CrossRef]
54. United Nations Transforming Our World: The 2030 Agenda for Sustainable Development. Available online: https://sdgs.un.org/2030agenda (accessed on 30 August 2022).
55. Li, C.; Wei, D.; Vause, J.; Liu, J. Towards a Societal Scale Environmental Sensing Network with Public Participation. *Int. J. Sustain. Dev. World Ecol.* **2013**, *20*, 261–266. [CrossRef]
56. Overeem, A.; Robinson, J.C.R.; Leijnse, H.; Steeneveld, G.J.; Horn, B.K.P.; Uijlenhoet, R. Crowdsourcing Urban Air Temperatures from Smartphone Battery Temperatures. *Geophys. Res. Lett.* **2013**, *40*, 4081–4085. [CrossRef]
57. Droste, A.M.; Pape, J.J.; Overeem, A.; Leijnse, H.; Steeneveld, G.J.; Van Delden, A.J.; Uijlenhoet, R. Crowdsourcing Urban Air Temperatures through Smartphone Battery Temperatures in São Paulo, Brazil. *J. Atmos. Ocean. Technol.* **2017**, *34*, 1853–1866. [CrossRef]
58. Fatehian, S.; Jelokhani-Niaraki, M.; Kakroodi, A.A.; Dero, Q.Y.; Samany, N.N. A Volunteered Geographic Information System for Managing Environmental Pollution of Coastal Zones: A Case Study in Nowshahr, Iran. *Ocean Coast. Manag.* **2018**, *163*, 54–65. [CrossRef]

59. Xu, S.; Chen, X.; Pi, X.; Joe-Wong, C.; Zhang, P.; Noh, H.Y. ILOCuS: Incentivizing Vehicle Mobility to Optimize Sensing Distribution in Crowd Sensing. *IEEE Trans. Mob. Comput.* **2020**, *19*, 1831–1847. [CrossRef]
60. Ren, Y.; Wang, T.; Zhang, S.; Zhang, J. An Intelligent Big Data Collection Technology Based on Micro Mobile Data Centers for Crowdsensing Vehicular Sensor Network. *Pers. Ubiquitous Comput.* **2020**, 1–7. [CrossRef]
61. Rahim, A.; Ma, K.; Zhao, W.; Tolba, A.; Al-Makhadmeh, Z.; Xia, F. Cooperative Data Forwarding Based on Crowdsourcing in Vehicular Social Networks. *Pervasive Mob. Comput.* **2018**, *51*, 43–55. [CrossRef]
62. Ruiz-Correa, S.; Santani, D.; Ramírez-Salazar, B.; Ruiz-Correa, I.; Rendón-Huerta, F.A.; Olmos-Carrillo, C.; Sandoval-Mexicano, B.C.; Arcos-Garcia, Á.H.; Hasimoto-Beltrán, R.; Gatica-Perez, D. SenseCityVity: Mobile Crowdsourcing, Urban Awareness, and Collective Action in Mexico. *IEEE Pervasive Comput.* **2017**, *16*, 44–53. [CrossRef]
63. Pateman, R.; Tuhkanen, H.; Cinderby, S. Citizen Science and the Sustainable Development Goals in Low and Middle Income Country Cities. *Sustain. Switz.* **2021**, *13*, 9534. [CrossRef]
64. Zhang, Y.; Li, X.; Wang, A.; Bao, T.; Tian, S. Density and Diversity of OpenStreetMap Road Networks in China. *J. Urban Manag.* **2015**, *4*, 135–146. [CrossRef]
65. Zhao, P.; Jia, T.; Qin, K.; Shan, J.; Jiao, C. Statistical Analysis on the Evolution of OpenStreetMap Road Networks in Beijing. *Phys. Stat. Mech. Appl.* **2015**, *420*, 59–72. [CrossRef]
66. Zia, M.; Cakir, Z.; Seker, D.Z. Turkey OpenStreetMap Dataset - Spatial Analysis of Development and Growth Proxies. *GeoScape* **2019**, *11*, 140–151. [CrossRef]
67. de Leeuw, J.; Said, M.; Ortegah, L.; Nagda, S.; Georgiadou, Y.; DeBlois, M. An Assessment of the Accuracy of Volunteered Road Map Production in Western Kenya. *Remote Sens.* **2011**, *3*, 247–256. [CrossRef]
68. Quinn, S. Using Small Cities to Understand the Crowd behind OpenStreetMap. *GeoJournal* **2017**, *82*, 455–473. [CrossRef]
69. Cilliers, L.; Flowerday, S. Factors That Influence the Usability of a Participatory IVR Crowdsourcing System in a Smart City. *S. Afr. Comput. J.* **2017**, *29*, 16–30. [CrossRef]
70. Bugs, G.; Granell, C.; Fonts, O.; Huerta, J.; Painho, M. An Assessment of Public Participation GIS and Web 2.0 Technologies in Urban Planning Practice in Canela, Brazil. *Cities* **2010**, *27*, 172–181. [CrossRef]
71. Orrego, R.; Barbosa, J. A Model for Resource Management in Smart Cities Based on Crowdsourcing and Gamification. *J. Univers. Comput. Sci.* **2019**, *25*, 1018–1038.
72. Wang, L.; Fan, H.; Wang, Y. Fine-Resolution Population Mapping from International Space Station Nighttime Photography and Multisource Social Sensing Data Based on Similarity Matching. *Remote Sens.* **2019**, *11*, 1900. [CrossRef]
73. Wang, L.; Fan, H.; Wang, Y. Improving Population Mapping Using Luojia 1-01 Nighttime Light Image and Location-Based Social Media Data. *Sci. Total Environ.* **2020**, *730*, 139148. [CrossRef] [PubMed]
74. Yao, Y.; Liu, X.; Li, X.; Zhang, J.; Liang, Z.; Mai, K.; Zhang, Y. Mapping Fine-Scale Population Distributions at the Building Level by Integrating Multisource Geospatial Big Data. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 1220–1244. [CrossRef]
75. Jing, C.; Zhou, W.; Qian, Y.; Yan, J. Mapping the Urban Population in Residential Neighborhoods by Integrating Remote Sensing and Crowdsourcing Data. *Remote Sens.* **2020**, *12*, 3235. [CrossRef]
76. Gebremedhin, E.T.; Basco-Carrera, L.; Jonoski, A.; Iliffe, M.; Winsemius, H. Crowdsourcing and Interactive Modelling for Urban Flood Management. *J. Flood Risk Manag.* **2020**, *13*, e12602. [CrossRef]
77. Hirata, E.; Giannotti, M.A.; Larocca, A.P.C.; Quintanilha, J.A. Flooding and Inundation Collaborative Mapping – Use of the Crowdmap/Ushahidi Platform in the City of Sao Paulo, Brazil. *J. Flood Risk Manag.* **2018**, *11*, S98–S109. [CrossRef]
78. Horita, F.E.A.; de Albuquerque, J.P.; Degrossi, L.C.; Mendiondo, E.M.; Ueyama, J. Development of a Spatial Decision Support System for Flood Risk Management in Brazil That Combines Volunteered Geographic Information with Wireless Sensor Networks. *Comput. Geosci.* **2015**, *80*, 84–94. [CrossRef]
79. Jelokhani-Niaraki, M.; Bastami Mofrad, R.; Yazdanpanah Dero, Q.; Hajiloo, F.; Sadeghi-Niaraki, A. A Volunteered Geographic Information System for Monitoring and Managing Urban Crimes: A Case Study of Tehran, Iran. *Police Pract. Res.* **2020**, *21*, 547–561. [CrossRef]
80. Bako, A.I.; Aduloju, O.T.B.; Osewa, D.J.; Anofi, A.O.; Abubakar-Karma, A.T. Application of Participatory GIS in Crime Mapping of Ibadan North, Nigeria. *Pap. Appl. Geogr.* **2021**, *7*, 183–198. [CrossRef]
81. Levy, P. *L'Intelligence Collective: Pour une Anthropologie du Cyberspace*; La Découverte Paris: Paris, France, 1994.
82. Elwood, S.; Goodchild, M.F.; Sui, D.Z. Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Ann. Assoc. Am. Geogr.* **2012**, *102*, 571–590. [CrossRef]
83. Parker, C.J.; May, A.; Mitchell, V. User-Centred Design of Neogeography: The Impact of Volunteered Geographic Information on Users' Perceptions of Online Map 'Mashups'. *Ergonomics* **2014**, *57*, 987–997. [CrossRef]
84. Miao, R.; Wang, Y.; Li, S. Analyzing Urban Spatial Patterns and Functional Zones Using Sina Weibo Poi Data: A Case Study of Beijing. *Sustainability* **2021**, *13*, 647. [CrossRef]
85. Morales, J.; Flacke, J.; Morales, J.; Zevenbergen, J. Mapping Urban Accessibility in Data Scarce Contexts Using Space Syntax and Location-Based Methods. *Appl. Spat. Anal. Policy* **2019**, *12*, 205–228. [CrossRef]
86. Devkota, B.; Miyazaki, H.; Witayangkurn, A.; Kim, S.M. Using Volunteered Geographic Information and Nighttime Light Remote Sensing Data to Identify Tourism Areas of Interest. *Sustainability* **2019**, *11*, 4718. [CrossRef]
87. Chang, S.; Wang, Z.; Mao, D.; Guan, K.; Jia, M.; Chen, C. Mapping the Essential Urban Land Use in Changchun by Applying Random Forest and Multi-Source Geospatial Data. *Remote Sens.* **2020**, *12*, 2488. [CrossRef]

88. Anderson, J.; Sarkar, D.; Palen, L. Corporate Editors in the Evolving Landscape of OpenStreetMap. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 232. [CrossRef]
89. Butt, M.A.; Li, S.; Javed, N. Towards Co-PPGIS—A Collaborative Public Participatory GIS-Based Measure for Transparency in Housing Schemes: A Case of Lahore, Pakistan. *Appl. Geomat.* **2016**, *8*, 27–40. [CrossRef]
90. Mansourian, A.; Taleai, M.; Fasihi, A. A Web-Based Spatial Decision Support System to Enhance Public Participation in Urban Planning Processes. *J. Spat. Sci.* **2011**, *56*, 269–282. [CrossRef]
91. Chatterjee, S.; Lim, S. A Multi-Objective Differential Evolutionary Method for Constrained Crowd Judgment Analysis. *IEEE Access* **2020**, *8*, 87647–87664. [CrossRef]
92. McArthur, D.P.; Hong, J. Visualising Where Commuting Cyclists Travel Using Crowdsourced Data. *J. Transp. Geogr.* **2019**, *74*, 233–241. [CrossRef]
93. Goletz, M.; Ehebrecht, D. How Can GPS/GNSS Tracking Data Be Used to Improve Our Understanding of Informal Transport? A Discussion Based on a Feasibility Study from Dar Es Salaam. *J. Transp. Geogr.* **2020**, *88*, 102305. [CrossRef]
94. Kemajou, A.; Jaligot, R.; Bosch, M.; Chenal, J. Assessing Motorcycle Taxi Activity in Cameroon Using GPS Devices. *J. Transp. Geogr.* **2019**, *79*, 102472. [CrossRef]
95. Misra, A.; Gooze, A.; Watkins, K.; Asad, M.; Le Dantec, C. Crowdsourcing and Its Application to Transportation Data Collection and Management. *Transp. Res. Rec.* **2014**, 1–8. [CrossRef]
96. Alhalabi, W.; Lytras, M.; Aljohani, N. Crowdsourcing Research for Social Insights into Smart Cities Applications and Services. *Sustain. Switz.* **2021**, *13*, 7531. [CrossRef]
97. Anik, M.A.H.; Sadeek, S.N.; Hossain, M.; Kabir, S. A Framework for Involving the Young Generation in Transportation Planning Using Social Media and Crowd Sourcing. *Transp. Policy* **2020**, *97*, 1–18. [CrossRef]
98. Belkahla Driss, O.; Mellouli, S.; Trabelsi, Z. From Citizens to Government Policy-Makers: Social Media Data Analysis. *Gov. Inf. Q.* **2019**, *36*, 560–570. [CrossRef]
99. Nelson, T.; Ferster, C.; Laberee, K.; Fuller, D.; Winters, M. Crowdsourced Data for Bicycling Research and Practice. *Transp. Rev.* **2021**, *41*, 97–114. [CrossRef]
100. Internet World Stats Internet World Penetration Rates by Geographic Regions. 2022. Available online: https://www.internetworldstats.com/stats.htm (accessed on 12 September 2022).
101. GSMA. The State of Mobile Internet Connectivity. 2020. Available online: https://www.gsma.com/r/wp-content/uploads/2020/09/GSMA-State-of-Mobile-Internet-Connectivity-Report-2020.pdf (accessed on 14 July 2021).
102. Young, J.C.; Lynch, R.; Boakye-Achampong, S.; Jowaisas, C.; Sam, J.; Norlander, B. Volunteer Geographic Information in the Global South: Barriers to Local Implementation of Mapping Projects across Africa. *GeoJournal* **2021**, *86*, 2227–2243. [CrossRef]
103. Omanga, D.; Mainye, P.C. North-South Collaborations as a Way of 'Not Knowing Africa': Researching Digital Technologies in Kenya. *J. Afr. Cult. Stud.* **2019**, *31*, 273–275. [CrossRef]
104. Young, J.C. The New Knowledge Politics of Digital Colonialism. *Environ. Plan. A* **2019**, *51*, 1424–1441. [CrossRef]
105. Lepenies, R.; Zakari, I.S. Citizen Science for Transformative Air Quality Policy in Germany and Niger. *Sustainability* **2021**, *13*, 3973. [CrossRef]
106. Jaligot, R.; Kemajou, A.; Chenal, J. Cultural Ecosystem Services Provision in Response to Urbanization in Cameroon. *Land Use Policy* **2018**, *79*, 641–649. [CrossRef]
107. Williams, S.; White, A.; Waiganjo, P.; Orwa, D.; Klopp, J. The Digital Matatu Project: Using Cell Phones to Create an Open Source Data for Nairobi's Semi-Formal Bus System. *J. Transp. Geogr.* **2015**, *49*, 39–51. [CrossRef]
108. Milusheva, S.; Marty, R.; Bedoya, G.; Williams, S.; Resor, E.; Legovini, A. Applying Machine Learning and Geolocation Techniques to Social Media Data (Twitter) to Develop a Resource for Urban Planning. *PLoS ONE* **2021**, *16*, e0244317. [CrossRef] [PubMed]
109. El Alaoui El Abdallaoui, H.; El Fazziki, A.; Ouarzazi, J.; Ennaji, F.Z.; Sadgal, M. A Crowdsensing-Based Framework for Urban Air Quality Decision Support. *Turk. J. Electr. Eng. Comput. Sci.* **2019**, *27*, 4298–4313. [CrossRef]
110. de Lima, A.C.L.; Gusmão, A.D.; de Menezes Cruz, M.L.P.; dos Santos, E.C.G. The Use of Information and Communication Technology for the Construction and Demolition Waste (CDW) Management in the City of Recife. *Electron. J. Geotech. Eng.* **2015**, *20*, 4997–5008.
111. Camara, G.S.; Camboim, S.P.; Bravo, J.V.M. Collaborative Emotional Mapping as a Tool for Urban Mobility Planning. *Bol. Cienc. Geod.* **2021**, *27*. [CrossRef]
112. World Population Review Global South Countries. 2022. Available online: https://worldpopulationreview.com/country-rankings/global-south-countries (accessed on 1 August 2022).
113. Rigolon, A.; Browning, M.; Lee, K.; Shin, S. Access to Urban Green Space in Cities of the Global South: A Systematic Literature Review. *Urban Sci.* **2018**, *2*, 67. [CrossRef]

sustainability

MDPI

Article

Does Cross-Border E-Commerce Promote Economic Growth? Empirical Research on China's Pilot Zones

Min Zhong *, Zengtao Wang and Xing Ge

School of Economics and Finance, Xi'an Jiaotong University, Xi'an 710061, China
* Correspondence: zm-marina2020@stu.xjtu.edu.cn

Abstract: Whether the construction of China's cross-border e-commerce (CBEC) comprehensive pilot zones can promote economic growth and social sustainable development is an important question worthy of discussion. This paper uses the difference-in-differences (DID) method to test the impact of the establishment of CBEC comprehensive pilot zones on economic growth and discusses the impact mechanism. The results are as follows. (1) The construction of CBEC comprehensive pilot zones can promote economic growth. After testing with parallel trend, placebo, and other robustness methods, the results are still valid. (2) The economic promotion effect of the construction of CBEC comprehensive pilot zones will be more evident in the coastal and eastern regions. The economic promotion effect of the first, second, and third batch of CBEC comprehensive pilot zones is clear. (3) The main ways that the construction of CBEC comprehensive pilot zones can facilitate economic growth are through urban digitalization, trade openness, and information service industry agglomeration.

Keywords: cross-border e-commerce; DID; economic growth

Citation: Zhong, M.; Wang, Z.; Ge, X. Does Cross-Border E-Commerce Promote Economic Growth? Empirical Research on China's Pilot Zones. *Sustainability* **2022**, *14*, 11032. https://doi.org/10.3390/su141711032

Academic Editors: Tan Yigitcanlar, Bo Xia, Jamile Sabatini Marques and Tatiana Tucunduva Philippi Cortese

Received: 1 August 2022
Accepted: 31 August 2022
Published: 4 September 2022

1. Introduction

Cross-border e-commerce (CBEC) is a new form of international trade in the internet era. Compared with traditional trade, CBEC has fewer trade links, low transaction costs, and fewer intermediary links [1,2], which gradually stimulates the growth of international trade and facilitates economic growth for all countries [3]. In recent years, China's CBEC has developed rapidly. In 2021, the scale of China's CBEC market was CNY 14.2 trillion, including CNY 11 trillion for exports and CNY 3.2 trillion for imports. China's CBEC transactions accounted for 36.32% of its total import and export value of CNY 39.1 trillion of goods trade in 2021, which means that the penetration rate of the CBEC industry still exceeded 35%. It can be predicted that in the future, with the continuous growth of the industry scale, the penetration rate of the CBEC industry will also continue to increase. In addition to the advantages of CBEC, the achievements of China's CBEC depend on the strong support of, and promotion by, the Chinese government. From 2014 to 2022, the Chinese government work report has mentioned CBEC for nine consecutive years, emphasizing the need to accelerate the development of CBEC and other new formats and models, while expanding the CBEC market has become an important part of the construction of China's new double circulation development pattern.

Although CBEC has many advantages over traditional trade, in the early stage of CBEC development in China, due to the novel mode, China's rules and regulations lagged behind the practice, resulting in serious obstacles with regard to logistics, payment, customs clearance, taxation, foreign exchange settlement, and other CBEC trade links. Among them, the factors restricting the development of CBEC mainly include the following: first, the low efficiency of customs clearance caused by the fragmentation of CBEC transactions. CBEC orders have the characteristics of small batches, high frequency, multiple categories, and low value, which manifests in a large number of express deliveries and parcels. This leads to cumbersome customs formalities for imported and exported goods, which not only consumes a lot of manpower and material resources but also greatly prolongs the customs clearance time and reduces customs clearance efficiency. Second, CBEC procurement is relatively decentralized, which gives rise to trade cost problems. A large number of enterprises purchase goods from individuals or self-employed households. The lack of documents and tickets forces many enterprises to conduct business using gray customs clearance. The increase in relevant costs seriously hinders the development and expansion of the CBEC industry. Third, CBEC enterprises face funding constraints due to their small scale. As the main CBEC participants are small and medium-sized enterprises (SMEs), the internal ability of these small and medium-sized enterprises to stabilize capital flow is generally poor, and the external financing constraints are high. In addition, it is difficult for some CBEC enterprises to obtain legal and effective purchase vouchers, meaning they are unable to obtain export tax rebates. These factors cause heavy financial constraints for SMEs engaged in CBEC.

To promote the healthy development of China's CBEC and to implement a higher level of opening up to the outside world, China launched the CBEC comprehensive pilot zone project in 2015. Exploring appropriate policy tools by establishing pilot areas and promoting them nationwide is a method often used by the Chinese government to promote progressive reform [4,5]. According to the traditional economic growth theory, institutional innovation is a decisive factor in economic growth [6]. The Chinese government attempted to solve the institutional problems faced in the development of the CBEC industry in the pilot area by establishing CBEC comprehensive pilot zones to achieve economic growth. The CBEC comprehensive pilot zones focus on facilitating the business process of CBEC, improving the information and digital construction of the city, promoting the agglomeration of the CBEC industry, and forming a complete supply chain, thus solving the institutional problems that have arisen in the development of the CBEC industry.

The Chinese government launched the first batch of CBEC comprehensive pilot zones in March 2015, including only one city, Hangzhou. In January 2016, China promoted and copied the experience of the Hangzhou CBEC comprehensive pilot zone and launched the second batch of CEBC comprehensive pilot zones in 12 cities, including Tianjin and Hefei. In July 2018, the Chinese government replicated and promoted the mature practices of the first and second batch of 13 CBEC comprehensive pilot zones in 12 aspects, formed by the "six systems and two platforms", to the whole country, launching the third batch of CBEC comprehensive pilot zones in 22 cities, including Beijing. To date, the Chinese government has launched six batches of CBEC pilot zones, covering 30 provinces, autonomous regions, and cities, and forming a pattern of land–sea interaction and east–west mutual assistance. However, due to the late establishment of the latter three batches and the limited availability of data, our paper only explores the implementation effect of the first three batches.

China's CBEC has developed rapidly, and the first batch of CBEC comprehensive pilot zones has been operating for over 7 years. Through an economic effect evaluation of China's CBEC pilot zone policy, our paper aims to answer the following three questions: (1) In the context of China's development of the digital economy, are there differences between pilot and non-pilot zones for CBEC development? In other words, is the pilot policy effective? (2) What are the factors and mechanisms that impact the effectiveness of the pilot zone policy? In other words, what is the key to the effective implementation of the pilot policy? (3) After the implementation of the pilot zone policy, are there obvious changes in the pilot zones? In other words, which cities have effective pilot zone policies? To conduct systematic and rigorous research, our paper proposes a time-varying difference-in-differences (DID) model to evaluate the effectiveness of the CBEC comprehensive pilot zone policy based on panel data of Chinese prefecture-level cities from 2011 to 2019. The DID model is a widely used policy evaluation method that evaluates the net impact before and after the implementation of a policy by eliminating the influence of individual heterogeneity differences and time change factors. On this basis, we analyze the differences in policy effects in eastern, central, and western China, the differences in policy effects in coastal and inland areas, as well as the differences in policy effects in the first, second, and third batches of CBEC pilot zones. In addition, the robustness of benchmark regression is tested from three aspects. Finally, our paper analyzes the impact mechanism of the CBEC pilot zone policy and proposes policy suggestions to promote the development of China's CBEC pilot zone.

The main contributions of our paper lie in both the theoretical and practical aspects. Specifically, we use the time-varying DID model to innovatively study the economic effects of the batch implementation of policies in China's CBEC comprehensive pilot zone. Most importantly, in addition to the effect of policy implementation in the pilot area at the national level, we also compare the effect of policies in different regions with different levels of economic development and different batches. At the same time, we discuss the cumulative and dynamic effects of the CBEC pilot zone policy. In addition, we analyze the path of policy impact from three perspectives: the level of urban digitalization, the degree of trade openness, and information service industry agglomeration. Finally, we present suggestions for promoting the implementation of the CBEC comprehensive pilot zone policy in China and around the world.

The rest of our paper is organized as follows: Section 2 contains a review of the literature; Section 3 discusses the methodology, including the theoretical mechanism and the empirical strategy; Section 4 analyzes the empirical results; Section 5 contains the discussion; and Section 6 comprises the conclusion and suggestions.

2. Literature Review

In recent years, CBEC has developed rapidly, and the transaction scale has increased significantly. In the past two years in particular, due to the global COVID-19 pandemic, the development of the traditional economy has been seriously hindered, while CBEC has grown and developed rapidly. It is estimated that by 2025, the revenue of global CBEC will increase from USD 250 billion to USD 350 billion [7].

Many scholars have begun to study how to better promote the development of CBEC. The research has found that, in addition to the basic driving factors, such as internet infrastructure, capital investment, convenient means of payment, per capita education level, and spillover effects of other countries [8], it is also important to break down the tax barriers, such as consumption tax and value added tax, for the effective development of cross-border online trade [9,10]. At the same time, many studies also believe that simplifying import and export process regulation and reducing the regulatory differences across regions are important steps to promote the growth of CBEC [11]. However, the above-mentioned studies employ logical deduction at the theoretical level and lack the support of empirical evidence.

In addition, many scholars have begun to study the economic effects of CBEC. Lendle and Vézina [2] found that CBEC platforms such as eBay can significantly increase a company's export probability and export scale. Ma et al. [3] found that CBEC significantly promoted China's import growth. Other scholars have focused on exploring the impact of CBEC on transaction costs [12] and labor productivity [13]. In general, although the existing literature examines the economic impact of CBEC on international trade, transaction costs, and productivity, there is a lack of discussion on the direct impact of CBEC on economic growth.

To promote the development of CBEC in China, the Chinese government has adopted a top-down model of CBEC comprehensive pilot zones. In this model, the policy objectives of China's pilot zones to facilitate the construction of CBEC cities are formulated by the national government, while specific policy tools are formulated and issued by local governments. The direct goal of the policy is to promote international trade, but the ultimate goal is to achieve economic development. Therefore, our paper focuses on the impact of the construction of CBEC comprehensive pilot zones on the economic development of Chinese cities.

In the early stages, some scholars focused on the construction of a single CBEC comprehensive pilot zone. Lu and Wang [14] studied the development status and future competition direction of the Dalian and Tianjin comprehensive pilot zones. Jin et al. [15] examined the impact of the establishment of CBEC comprehensive pilot zones by analyzing the pilot zone in Henan Province from five aspects: transaction scale, growth space, industrial penetration, supporting environment, and platform agglomeration. With the gradual increase in CBEC comprehensive pilot zones, some scholars began to study the overall economic results of the pilot zones. Chen [16] found that the establishment of comprehensive pilot zones is closely related to economic growth by using the data of 35 CBEC pilot comprehensive zones, based on the gray correlation theory. Wang et al. [17] found that the CBEC comprehensive pilot zone policy has an important positive impact on the economic development of the pilot city by using the data of the first three batches of CBEC comprehensive pilot zones in China, based on the structural equation model.

In addition, some scholars have tested the effect of a policy tool through the synthetic control method (SCM) [18,19]. However, the SCM is usually suitable for comparative case studies, as the sample size for this method is small [20]. In addition, matching pilot cities by weight in SCM will lead to errors [21]. In contrast, the DID method regards the implementation of the policy in the pilot area as an independent variable, which not only limits the interaction between the independent variable and the dependent variable but also avoids the common endogenous problem in panel regression. In addition, because the sample grouping of cities in the pilot area is independent of individual heterogeneity, DID can control the impact of unobservable individual heterogeneity on dependent variables and extract the actual policy effect of the pilot area policy. Therefore, our paper focuses on the impact of the construction of CBEC comprehensive pilot zones on economic development, thus obtaining the "net effect" of the policy more effectively. However, considering the phased launch and implementation characteristics of China's CBEC comprehensive pilot zone policy, we use the time-varying DID model to more accurately analyze the effect of the policy.

In summary, recent research has mainly focused on the impact of CBEC on international trade but ignored economic growth. Furthermore, few people have explored the influence mechanism of CBEC on economic growth. Moreover, there is a lack of in-depth research on the heterogeneity between cities in the literature. In addition, most previous studies have provided static results rather than the possible dynamic impact of CBEC pilot zones. Therefore, the time-varying DID model is used to more accurately analyze the effect of the policy and the cumulative effect of the policy over time, as well as the different impacts on different regions of China.

3. Methodology

3.1. Theoretical Mechanism

The CBEC comprehensive pilot zone is a pilot economic zone with urban areas as the core, established with the support of national pilot policies. The CBEC comprehensive pilot zone approved by the Chinese government specifically refers to the adoption of specific management modes and special policy preferences for CBEC activities in some regions to achieve a more convenient CBEC. We combed the relevant documents approved by the Chinese government to unify the establishment of CBEC comprehensive pilot zones. It was found that the construction of CBEC comprehensive pilot zones mainly includes the following aspects.

First, each pilot zone focuses on innovating and improving the digital construction of the technical standards, business processes, regulatory models, and other CBEC transaction processes. Digital construction not only refers to the construction of information infrastructure but also includes the maintenance and updating of digital platforms, the realization of digital customs clearance, and the construction of online industrial parks.

Second, each pilot zone strives to improve trade facilitation in the CBEC comprehensive pilot zone. This is mainly reflected in the following aspects: first, in terms of tax supervision, the comprehensive pilot zone implemented the policy of "no ticket tax exemption", so that the CBEC enterprises in the comprehensive pilot zone can enjoy the preferential treatment of exemption from value added tax and consumption tax; and second, in terms of import and export declaration and customs clearance, the process has been greatly simplified. The General Administration of Customs, the General Administration of Taxation, and other government departments have issued a series of policies to help improve the efficiency of customs clearance, such as simplifying the classification of import and export goods, and handling the import and export procedures of CBEC goods through a single window. It has been reported that, under the above measures, the import and export declaration time for goods in the Hangzhou CBEC comprehensive pilot zone was shortened from 4 h to an average of 1 min.

Third, each pilot zone focuses on encouraging information service enterprises to gather. The comprehensive pilot zone generally carries out the construction of CBEC industrial parks, encouraging well-known e-commerce platform enterprises, CBEC upstream and downstream enterprises, and related service enterprises to settle in, and it provides comprehensive supply chain services, such as finance, customs clearance, quarantine, logistics, and talent in the park, to create a CBEC industrial ecosystem and to promote the overall growth and strength of the enterprises in the comprehensive pilot zone.

Based on the development background of the above comprehensive pilot zone, we believe that a CBEC comprehensive pilot zone will eventually achieve regional economic growth through the construction of urban digitization, an improvement in trade openness, and the realization of information service industry agglomeration. Therefore, based on the above analysis, we first construct a theoretical mechanism diagram (see Figure 1) of CBEC comprehensive pilot zone construction and the effects on regional economic growth, and then analyze these three mechanisms.

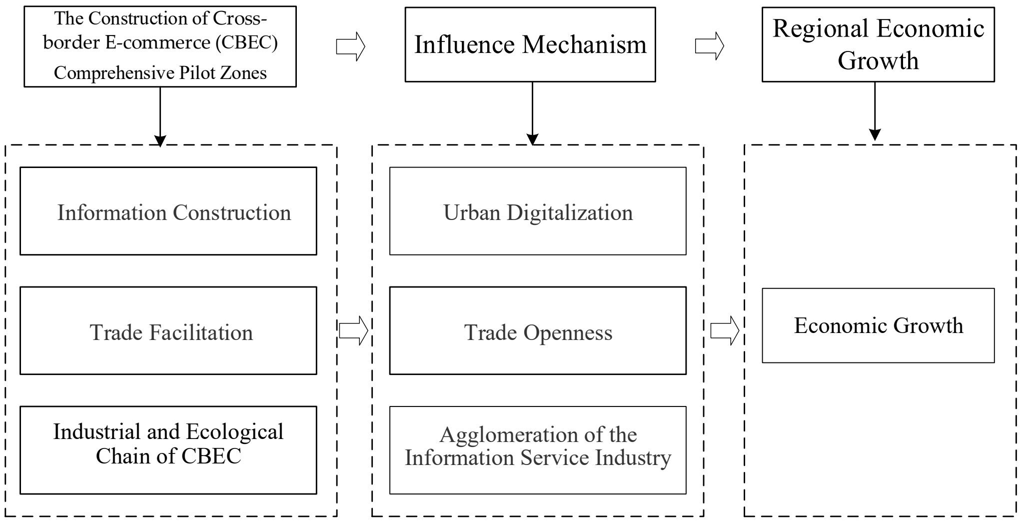


Figure 1. The construction of CBEC comprehensive pilot zones and their effect on regional economic growth.

First, the policies of CBEC comprehensive pilot zones impact economic growth by affecting urban digital construction. The policies of CBEC comprehensive pilot zones focus on information construction to promote the digital transformation of cities, and the digital transformation of cities contributes to economic growth [22,23].

Second, CBEC comprehensive pilot zones influence economic growth by affecting the degree of trade openness. CBEC comprehensive pilot zones focus on building the growth of CBEC, a new trade format, to create new competitive advantages in international trade and to finally achieve trade promotion effects.

Third, CBEC comprehensive pilot zones have an effect on economic growth by facilitating the agglomeration of the information service industry. An important task of the pilot zones is to gather CBEC industries, create a complete CBEC industrial chain and ecological chain, and finally achieve economic growth. At present, the industrial agglomeration of China's CBEC comprehensive pilot zones mainly reflects the agglomeration of online integrated service industries and offline industrial parks.

3.2. Empirical Strategy

3.2.1. Specification

The purpose of our study is to evaluate the effect of CBEC comprehensive pilot zones on promoting urban economic growth in China. The implementation of CBEC comprehensive pilot zones is regarded as a "quasi-natural experiment". Specifically, a difference-in-differences model is applied to empirical research, in which the first difference is city, and the second difference is time. Therefore, in our study, the analysis of the DID model will focus on comparing the differences in economic growth between the cities in pilot areas and the cities in non-pilot areas before and after the implementation of the policy. Specifically, following Liu and Qiu [24], as well as Pierce and Schott [25], the specific model is as follows:

$$\mathrm{GDPP}_{it} = \alpha + \beta_1 \mathrm{D}_{it} + \delta \mathrm{X}_{it} + \lambda_\mathrm{i} + \lambda_\mathrm{t} + \varepsilon_{it} \quad (1)$$

where i and t represent city and time; GDPP_{it} denotes the per capita GDP of city i in year t; λ_i is a time city-related fixed effect, which controls all the city-related factors that do not change with time; λ_t refers to the time-related fixed effect, which controls all the time-related factors that do not change with the city; X_{it} stands for a set of time-varying city-level variables; and ε_{it} is the error term. Following Bertrand et al. [26], we cluster the samples to the urban level because of the potential heteroscedasticity and serial autocorrelation.

D_{it} is the setting variable of the CBEC comprehensive pilot zone. If city i is a pilot city in year t, then $\mathrm{D}_{it} = 1$; otherwise, $\mathrm{D}_{it} = 0$. Thus, D_{it} is the core explanatory variable in our paper, and β_1 is the core estimation coefficient, which is the average treatment effect of the establishment of CBEC pilot zones on economic growth. If the estimation coefficient β_1 is significantly positive, it means that the establishment of CBEC comprehensive pilot zones has a positive impact on economic growth. If the estimation coefficient β_1 is significantly negative, it means that the establishment of CBEC comprehensive pilot zones has a negative impact on economic growth.

To further test the mechanism of the establishment of comprehensive pilot zones affecting economic growth, the mediating effect model is used to analyze the influence mechanism of the pilot zones on economic growth [27]. The basic assumption of this model is that explanatory variables affect explained variables through mediating variables. Specifically, we build Equation (2) based on Equation (1) to determine whether there is a mediating effect. If β_2 is not significant, then the comprehensive pilot zone policy has no significant impact on the mediating variables, so the analysis is terminated. Otherwise, we then build Equation (3):

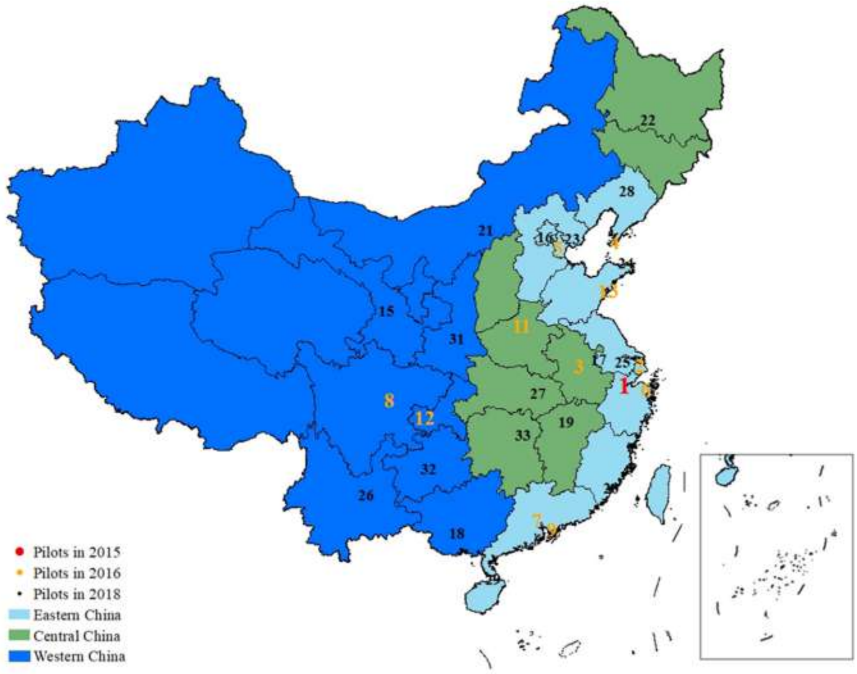
$$\mathrm{M}_{it} = \alpha + \beta_2 \mathrm{D}_{it} + \delta \mathrm{X}_{it} + \lambda_\mathrm{i} + \lambda_\mathrm{t} + \varepsilon_{it} \quad (2)$$

$$\mathrm{GDPP}_{it} = \alpha + \beta_3 \mathrm{D}_{it} + \tau \mathrm{M} + \delta \mathrm{X}_{it} + \lambda_\mathrm{i} + \lambda_\mathrm{t} + \varepsilon_{it} \quad (3)$$

In Equation (3), if τ is not significant, M has no mediating effect. Otherwise, if the coefficient is significant, M has a mediating effect. After introducing the mediating variable M, if β_3 is not significant in Equation (3), M is the only confirmed mediating variable. In other words, the influence path of the comprehensive pilot zone policy is unique and certain. Otherwise, there are other mediating variables or other influencing ways. In Equations (2) and (3), the mediating variables M include the degree of urban digitalization, trade openness, and the agglomeration level of the information service industry.

3.2.2. Data

Our paper selects 33 CBEC comprehensive pilot zones launched in 2015, 2016, and 2018 as the analysis object. The details of the cities, including the time and location of the pilots, are shown in Figure 2. In our study, cities are divided into eastern, central, and western regions according to the division of the National Development and Reform Commission of China, rather than geographical concepts.



Notes

1. Hangzhou	2. Shanghai	3. Hefei	4. Dalian	5. Tianjin	6. Ningbo
7. Guangzhou	8. Chengdu	9. Shenzhen	10. Suzhou	11. Zhengzhou	12. Chongqing
13. Qingdao	14. Dongguan	15. Lanzhou	16. Beijing	17. Nanjing	18. Nanning
19. Nanchang	20. Xiamen	21. Hohhot	22. Harbin	23. Tangshan	24. Weihai
25. Wuxi	26. Kunming	27. Wuhan	28. Shenyang	29. Haikou	30. Zhuhai
31. Xi'an	32. Guiyang	33. Changsha			

Figure 2. Details of pilot cities in 2015, 2016, and 2018.

There are 19 pilot zones in the eastern region, 6 pilot zones in the central region, and 8 pilot zones in the western region. The first two batches of pilot cities are mainly concentrated in the eastern region. The number of CBEC pilot zones in the eastern region is significantly higher than that in the central and western regions because the basic conditions for CBEC development in the central and western regions are relatively weak (for example, there are fewer e-commerce enterprises and fewer e-commerce talent resources). As a result of this layout, the development gap between the CBEC pilot zones in different regions will become clearer, showing that "the CBEC development in the eastern region will become stronger and stronger, while CBEC in the central and western regions will become weaker and weaker". To reverse this situation, the third batch of CBEC comprehensive pilot zones began to lean toward central and western cities.

In addition, after excluding the city samples with serious data loss, our paper uses the panel data of 117 cities from 2011 to 2019, including 33 cities in the treatment group and 84 cities in the control group. All data in our paper were obtained from the *China Urban Statistical Yearbook*. The selection and explanation of each variable are shown in Table 1.

Table 1. Definitions of variables.

Variables	Definition	Observation	Mean	Standard Deviation	Minimum Value	Maximum Value
GDPP	Real per capita GDP	1053	0.68	0.355	0.1162	2.183
D_{it}	CBEC pilot zones	1053	0.08	0.257	0	1
investment	Social fixed asset investment	1053	0.75	0.271	0.0036	1.798
consumption	Social consumption level	1053	3.25	2.535	0.3744	18.499
humancap	Human capital level	1053	4.66	1.076	2.241	6.798
roadper	Per capita road area	1053	0.02	0.018	0.000	0.105
fdi	Foreign direct investment level	1053	7.09	8.217	0.494	73.044
unemployment	Unemployment rate	1053	0.55	0.293	0.061	1.919
digitalscore	The level of urban digitalization	1053	0.02	0.070	0	0.552
trade	Trade openness	1053	0.77	2.604	0	11.411
agglomeration	Agglomeration level of the information service industry	1053	0.11	0.422	0	4.039

3.2.3. Explained Variables

We choose real per capita GDP (GDPP) to represent economic growth. Real per capita GDP is equal to the real GDP of the year divided by the total population at the end of the year. Specifically

$$GDPP_{it} = \frac{\text{Real GDP}_{it}}{\text{Total population}_{it}} \tag{4}$$

In Equation (4), the real GDP is calculated by using the nominal GDP and the deflator.

3.2.4. Core Explanatory Variables

Our paper examines whether CBEC comprehensive pilot zone policies play an important role in economic growth. D_{it} is the core explanatory variable of whether a city is a CBEC pilot city and when it becomes a CBEC pilot city. If city i is a pilot city in year t, then $D_{it}= 1$; otherwise, $D_{it}= 0$.

3.2.5. Control Variables

We also controlled other factors that may affect economic growth, such as social fixed asset investment (investment), social consumption level (consumption), human capital level (humancap), per capita road area (roadper), foreign direct investment level (fdi), and unemployment rate (unemployment). Specifically, investment is equal to total social fixed asset investment divided by GDP; consumption is equal to total retail sales of consumer goods divided by GDP; humancap is equal to taking the logarithm of the number of university students; roadper is equal to urban road area divided by the total population at the end of the year; fdi is equal to total foreign direct investment divided by GDP; and unemployment is equal to the proportion of unemployment in the total labor force.

3.2.6. Mediating Variables

The mediating variables we chose are: the level of urban digitalization (digitalscore), the degree of trade openness (trade), and the agglomeration level of the information service industry (agglomeration). The details are as follows.

First, we use the entropy weight method to aggregate several variables that can characterize the city's digitalization level into one variable, digitalscore. Specifically, several of the variables that reflect the level of urban digitalization are: the number of internet users per 100 people; the number of mobile phone users per 100 people; the proportion of computer services and software practitioners; per capita telecommunications business volume; per capita postal business volume; and digital inclusive finance index.

Second, the degree of trade openness (trade) is equal to the logarithm of the city's import and export trade volume.

Third, regarding the measurement of the agglomeration level of the information service industry (agglomeration), following O'Donoghue and Gleave [28] and Freedman [29], we use the location entropy method to measure agglomeration. Specifically, the calculation formula of the agglomeration level of *r* industry in city *i* during the period of *t* is

$$\text{agglomeration}_{irt} = \left(\text{e}_{irt} / \sum_{i} \text{e}_{irt}\right) / \left(\sum_{i} \text{e}_{irt} / \sum_{i}\sum_{r} \text{e}_{irt}\right) \quad (5)$$

where $\text{agglomeration}_{irt}$ is the agglomeration degree of *r* industry in city *i* during period *t*; and e_{irt} refers to the number of employees of *r* industry in city *i* during period *t*. The larger the $\text{agglomeration}_{irt}$, the more the *r* industry in the city is clustered. In our paper, *r* refers to the information service industry.

4. Empirical Analyses

4.1. Benchmark Regression Results

Table 2 presents the main results for the DID specification. Column (1) only includes the pilot implementation of CBEC comprehensive pilot zones as an explanatory variable. Column (1) shows that the estimated coefficient of D_{it} is 0.568, which is significantly positive at the 1% level, indicating that, compared with other cities that are not CBEC comprehensive pilot zones, the establishment of CBEC comprehensive pilot zones will enable the per capita real GDP of the cities where the comprehensive pilot zones are located to increase by 56.8%. In column (2), we added city-level control variables. Column (2) shows that the estimated coefficient of D_{it} is 0.363, which is significantly positive at the 1% level. Column (3) adds more year-level control variables, and our regressor of interest, D_{it}, is statistically significant and positive at the 1% level.

Table 2. Baseline results.

	(1)	(2)	(3)	(4)
D_{it}	0.568 *** (0.09)	0.363 *** (0.03)	0.125 *** (0.03)	0.104 *** (0.03)
investment				0.067 * (0.04)
consumption				0.054 *** (0.01)
humancap				−0.028 (0.04)
roadper				−0.026 (0.54)
fdi				0.002 (0.00)
unemployment				−0.042 (0.04)
Constant	0.635 *** (0.03)	0.651 *** (0.00)	0.669 *** (0.00)	0.584 *** (0.20)
City fixed effect	no	yes	yes	yes
Year fixed effect	no	no	yes	yes
N	1053	1053	1053	1053
R^2	0.169	0.863	0.970	0.975

Note: Robust standard errors clustered at city in parentheses. *** Significant at the 1% level. * Significant at the 10% level.

Column (4) adds more control variables. Column (4) suggests that the coefficient of D_{it} is 0.104, which is significantly positive at the 1% level, indicating that CBEC comprehensive pilot zones can effectively stimulate local economic growth by 10.4%.

In column (4), the estimated coefficient of investment is significantly positive, which means that for every 1% increase in social fixed asset investment, the local per capita GDP will increase by 6.7%. The estimated coefficient of consumption is significantly positive, indicating that for every 1% increase in social consumption, the local per capita GDP will increase by 5.4%. The impact of other control variables on per capita GDP is not significant and has no statistical significance, but if these variables are discarded, it may result in missing variables.

Therefore, Table 2 implies that CBEC comprehensive pilot zone implementation can significantly promote the growth of local per capita GDP and improve the welfare of residents.

4.2. Parallel Trend Test and Dynamic Effect Analysis

The premise of using the DID method is to meet the parallel trend assumption. In other words, when it is not impacted by the pilot policy, the per capita GDP level of the treatment group and the control group has the same change trend. Due to the different times the pilot cities are impacted by the policy, we cannot simply set the virtual variable of time as the critical point of policy occurrence in a certain year; we need to set the virtual variable of the relative time value of policy implementation in the CBEC comprehensive pilot zone for each city time point. At the same time, because the CBEC pilot is affected by factors such as the intensity of policy implementation, the basis of policy implementation, and the adjustment of production factors, the policy effect of the CBEC pilot may have a buffer period, resulting in a certain delay in the implementation effect of the policy. Based on these two considerations, we follow Beck et al. [30] and use the event study method to build the following dynamic model:

$$\mathrm{GDPP}_{it} = \alpha + \sum_{k \geq -4}^{4} \beta_k \mathrm{D}_{it}^{k} + \delta \mathrm{X}_{it} + \lambda_i + \lambda_\mathrm{t} + \varepsilon_{it} \quad (6)$$

where i and t represent city and year, respectively; GDPP_{it} denotes the actual per capita GDP; and D_{it}^{k} indicates that the event of the establishment of CBEC pilot zones is a dummy variable. The assignment rules of D_{it}^{k} are as follows.

D_{it}^{k} is the relative time dummy variable. Provided that the year when city i became a CBEC pilot city is s (s = 2015, 2016, 2018), then we set $k = t–s$. When k is negative, if t is smaller than the year when the CBEC pilot policy is implemented, then we set $\mathrm{D}_{it}^{k} = 1$; otherwise, we set $\mathrm{D}_{it}^{k} = 0$. When k is no smaller than 0, if t is larger than the year when the CBEC pilot policy is implemented, then we set $\mathrm{D}_{it}^{k} = 1$; otherwise, we set $\mathrm{D}_{it}^{k} = 0$. In Equation (6), the year before the pilot establishment of CBEC is taken as the benchmark year.

The results are shown in Figure 3. Before the policy, the relative time dummy variable coefficients are not significant, and the values are small, which shows that, before the policy, there is no significant difference between the treatment group and the control group in the per capita GDP level. In other words, the CBEC pilot zone policy conforms to the parallel trend hypothesis. In terms of the dynamic effects of the policy, considering that, as of 2019, the first batch of pilot cities have had the policy in place for nearly five years, but the first batch is only Hangzhou, our paper mainly analyzes the dynamic effects in the four periods. The results show that two years after the implementation of the pilot policy, the impact coefficient of the CBEC pilot zone policy is significantly positive and rising, indicating that the CBEC pilot zone policy can produce the policy effect of stimulating the per capita GDP; however, it has a certain lag.

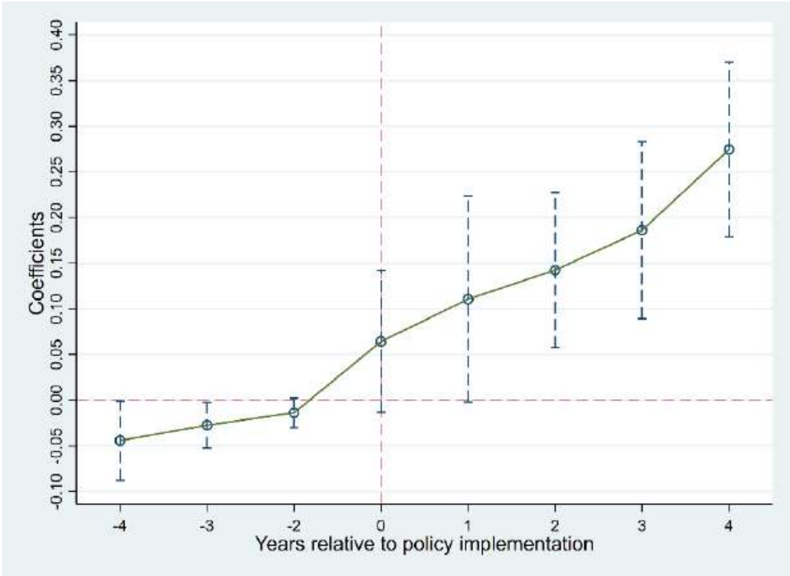


Figure 3. Parallel trend test. Note: The upper and lower dotted lines of the hollow points represent the 95% confidence interval.

4.3. Robustness Test

4.3.1. Changing City Samples

Considering the limited number of urban samples for the benchmark regression in this paper, only the first-, second-, third-tier, and new first-tier cities in China are selected. Therefore, we added more data to the original base, increasing the number of urban samples to 266. Table 3 shows a comparison between the estimated results and the benchmark regression of Table 2. The D_{it} result is still significant, but the estimated coefficient is increased. This shows that after considering more small cities as the control group, the economic effect of CBEC pilot zones will be clearer.

Table 3. Regression results of changing city samples.

	(1)	(2)	(3)	(4)
D_{it}	0.689 *** (0.09)	0.363 *** (0.03)	0.177 *** (0.03)	0.124 *** (0.03)
investment				0.045 *** (0.01)
consumption				0.069 *** (0.01)
humancap				0.016 (0.03)
roadper				0.207 (0.13)
fdi				0.008 ** (0.00)
unemployment				−0.020 (0.02)
Constant	0.521 *** (0.02)	0.532 *** (0.00)	0.539 *** (0.00)	0.260 ** (0.13)
City fixed effect	no	yes	yes	yes
Year fixed effect	no	no	yes	yes
N	2394	2394	2394	2394
R^2	0.121	0.893	0.968	0.976

Note: Robust standard errors clustered at city in parentheses. *** Significant at the 1% level. ** Significant at the 5% level.

4.3.2. Changing the Fourth and Fifth Batch Groups of Pilot Cities

In the benchmark regression, as the sample time does not include 2020, the fourth and fifth batches of pilot cities are included in the regression as control groups. Considering that the samples of the fourth and fifth batches of pilot cities may affect the estimated results, we use two methods to check the robustness of the benchmark regression results. (1) We extend the time of the sample to 2020 and introduce the fourth and fifth batches of pilot cities into the sample as treatment groups. The estimated results are shown in columns (1)–(4) of Table 4. Compared with Table 2, the estimated coefficient of D_{it} is significant at the level of 1%. (2) To ensure that the policy effect of the CBEC pilot will not be affected by the fourth and fifth batches during the sample period, we remove the samples of the fourth and fifth batches and re-estimate them. The estimated results are shown in columns (5)–(8) of Table 4. Compared with Table 2, in columns (5)–(8), the estimated coefficient of D_{it} is significant at the level of 1%.

Table 4. Regression results of changing the fourth and fifth batch groups of pilot cities.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D_{it}	0.509 *** (0.06)	0.350 *** (0.02)	0.116 *** (0.02)	0.100 *** (0.03)	0.530 *** (0.08)	0.360 *** (0.03)	0.131 *** (0.03)	0.114 *** (0.03)
investment				0.057 (0.04)				0.098 * (0.05)
consumption				0.038 *** (0.01)				0.045 *** (0.01)
humancap				−0.028 (0.04)				−0.059 (0.05)
roadper				−0.213 (0.60)				−0.255 (0.75)
fdi				0.003 (0.00)				0.001 (0.01)
unemployment				−0.051 (0.04)				−0.049 (0.05)
Constant	0.638 *** (0.03)	0.658 *** (0.00)	0.688 *** (0.00)	0.666 *** (0.20)	0.666 *** (0.04)	0.687 *** (0.00)	0.716 *** (0.00)	0.786 *** (0.24)
City fixed effect	no	yes	yes	yes	no	yes	yes	yes
Year fixed effect	no	no	yes	yes	no	no	yes	yes
N	1170	1170	1170	1170	648	648	648	630
R^2	0.192	0.870	0.966	0.970	0.173	0.902	0.971	0.974

Note: Robust standard errors clustered at city in parentheses. *** Significant at the 1% level. * Significant at the 10% level.

4.3.3. Placebo Test

Following La et al. [31], Abadie et al. [32], and Ma et al. [3], we divided the placebo test into two steps. First, 33 urban samples were randomly selected as the treatment group, while the rest of the cities were used as the control group. Second, we set the CBEC pilot city establishment time at random. Figure 4 shows the distribution of the estimates from 500 runs along with the baseline results. The distribution of the estimates from the random assignments is centered around zero, and the standard deviation of the estimates is not significant. The dotted line in Figure 4 represents the estimated coefficient of D_{it}, which is significantly different from the estimated value of the coefficient obtained in the placebo test, thus confirming that the effect of CBEC pilot city establishment on improving GDP does not come from unobservable factors.

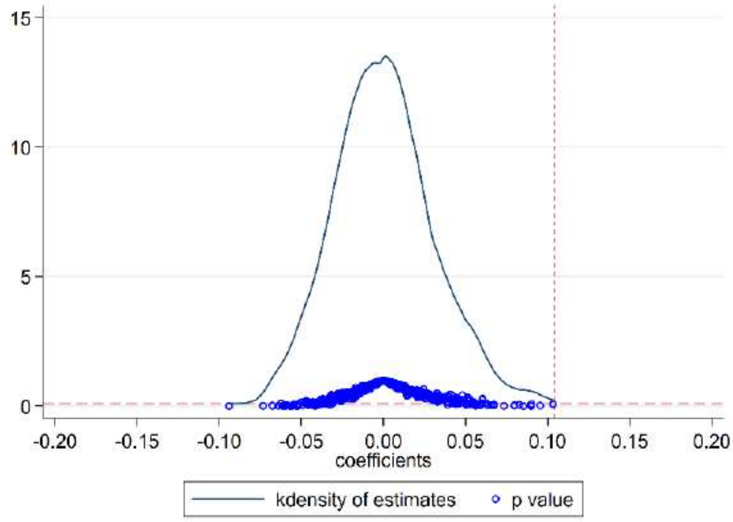


Figure 4. Placebo test.

4.4. Heterogeneity Analysis

4.4.1. Differences between Cities in Coastal and Inland Areas

As shown in column (1) of Table 5, the estimated coefficient of D_{it} in coastal areas is 0.135, which is significant at the level of 1%. As shown in column (2) of Table 5, the estimated coefficient of D_{it} in inland areas is 0.056, which is significant at the level of 10%. This empirical result shows that a CBEC pilot zone will promote the economic growth of coastal areas more evidently than that of inland areas. One possible reason for this is that, compared with inland areas, coastal areas have the advantages of resource endowments, such as commodity logistics network systems, CBEC talents, and regional advantages, which provide more assistance for the development of CBEC. Therefore, it will be easier for the construction of CBEC pilot zones to achieve good results.

Table 5. Regression results of heterogeneity analysis.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Cities in Coastal Areas	**Cities in Inland Areas**	**East**	**Middle**	**West**	**The First Batch**	**The Second Batch**	**The Third Batch**
D_{it}	0.135 ***	0.056 *	0.126 ***	0.059	0.013	0.163 ***	0.113 ***	0.131 ***
	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
investment	0.039	0.086 *	0.037	0.014	0.153*	0.178 **	0.176 **	0.102 *
	(0.07)	(0.04)	(0.05)	(0.05)	(0.08)	(0.08)	(0.07)	(0.06)
consumption	0.033 **	0.078 ***	0.044 ***	0.114 ***	0.033*	0.065 ***	0.073 ***	0.048 ***
	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
humancap	−0.025	−0.012	−0.005	0.062	−0.210	0.008	−0.015	−0.061
	(0.06)	(0.04)	(0.04)	(0.07)	(0.12)	(0.03)	(0.04)	(0.04)
roadper	1.570	−0.844	0.807	1.894 **	−2.060	−2.910 **	−1.487	−0.975
	(0.95)	(0.72)	(0.60)	(0.70)	(1.43)	(1.29)	(1.02)	(0.90)
fdi	0.001	0.006 *	0.002	0.007	0.007 *	0.016 *	−0.016	0.007 ***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)
unemployment	−0.083	−0.018	−0.131 **	0.011	0.072	0.066	0.043	−0.038
	(0.07)	(0.04)	(0.05)	(0.04)	(0.08)	(0.07)	(0.06)	(0.06)
Constant	0.696 **	0.395 *	0.607 ***	−0.125	1.432 **	0.204	0.456 **	0.698 ***
	(0.27)	(0.22)	(0.20)	(0.31)	(0.62)	(0.14)	(0.19)	(0.24)
City fixed effect	yes	yes	yes	yes	yes	yes	yes	yes
Year fixed effect	yes	yes	yes	yes	yes	yes	yes	yes
N	351	702	639	261	153	342	441	513
R^2	0.975	0.977	0.975	0.985	0.983	0.980	0.978	0.974

Note: Robust standard errors clustered at city in parentheses. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

4.4.2. Differences between Eastern, Central, and Western China

CBEC pilot cities are widely distributed in various regions of eastern, central, and western China, providing conditions for us to analyze the regional heterogeneity of the policy effects. Columns (3) to (5) of Table 5 show the economic impact of CBEC pilot policies on eastern, central, and western cities. The estimated coefficient of D_{it} in the eastern region is significantly positive, while the estimated coefficient of D_{it} in the central and western regions fails to pass the significance test. This shows that the establishment of CBEC pilot cities has not achieved positive results in the central and western regions, where the construction of CBEC development infrastructure software and hardware is relatively weak, and CBEC talents are scarce. As a result of this impact, the "siphon effect" of the eastern region is becoming increasingly significant. Finally, the development of CBEC pilot cities in the eastern, central, and western regions shows the "Matthew effect", in which the eastern region is an example of "the strong get stronger" and the central and western regions display "the weak get weaker" effect. To break this situation, to narrow the development gap between the east and the west, and to achieve leapfrog development for China's CBEC, they need to work together.

4.4.3. Differences between Different Batches

Columns (6) to (8) of Table 5 show the economic effects of different batches of CBEC pilot zones. According to the estimation coefficient of D_{it}, the impact of the first three batches of CBEC pilot zone policies on economic growth is significantly positive, which shows that each batch of CBEC pilot zone policy has achieved good economic results. Our further analysis found that the estimated coefficient of the first batch of CBEC pilot policies is higher than that of the third batch of CBEC pilot policies, while the estimated coefficient of the third batch of CBEC pilot policies is higher than that of the second batch of CBEC pilot policies. This shows that there is a certain relationship between the approval time of a CBEC pilot zone and its economic performance, but it is not an absolutely positive relationship. One possible reason for this is that the operation effect of a CBEC pilot zone is closely related to the talent, regional advantages, and other resource endowments of a city's CBEC development.

4.5. Influencing Mechanism Analysis

Column (1) of Table 6 shows that the estimated coefficient of D_{it} is 0.041, which is significant at the level of 1%. As shown in column (4), the estimated coefficients of D_{it} and digitalscore for economic growth are positive at the significance level of 1%. This empirical result means that D_{it} can improve economic growth by enabling the improvement of urban digital construction. This is because digital construction is an important part of the establishment of CBEC pilot zones, and digital technology helps promote economic growth [33,34]. Therefore, urban digital construction is a mediating mechanism for the construction of CBEC pilot zones to promote economic growth.

Column (2) of Table 6 reveals that the estimated coefficient of D_{it} is 0.047, which is significant at the level of 10%. As shown in column (5), the estimated coefficients of D_{it} and trade for economic growth are positive at the significance level of 1%. This means that D_{it} can improve economic growth by enabling a degree of trade openness in cities. This is because CBEC is a new model of international trade rooted in internet information technology [3]. Therefore, one of the most important ways for CBEC pilot zone construction to stimulate urban economic growth is to enable a degree of trade openness.

Column (3) of Table 6 shows that the estimated coefficient of D_{it} is 0.338, which is significant at the level of 1%. As shown in column (6), the estimated coefficients of D_{it} and agglomeration for economic growth are significantly positive. This shows that the construction of a CBEC pilot zone can promote an improvement in local GDP by gathering information service industries. One possible reason for this is that one of the most important tasks in building a CBEC pilot zone is to build a complete CBEC industrial chain and ecological chain, and to form an industrial agglomeration of online integrated services and offline industrial parks.

To sum up, the above results show that the construction of a CBEC pilot zone affects economic growth mainly through three channels: urban digitalization, trade openness, and the agglomeration of the information service industry. In other words, the construction of a CBEC pilot zone promotes economic growth by improving the level of urban digitalization, the degree of trade openness, and the agglomeration level of the information service industry.

Table 6. Regression results of influence mechanism.

	(1)	(2)	(3)	(4)	(5)	(6)
	Digitalscore	**Trade**	**Agglomeration**	**GDPP**	**GDPP**	**GDPP**
D_{it}	0.041 *** (0.01)	0.047 * (0.03)	0.338 *** (0.07)	0.090 *** (0.03)	0.095 *** (0.03)	0.095 *** (0.03)
digitalscore				0.355 ** (0.15)		
trade					0.207 ** (0.10)	
agglomeration						0.029 ** (0.01)
Other variables	yes	yes	yes	yes	yes	yes
City fixed effect	yes	yes	yes	yes	yes	yes
Year fixed effect	yes	yes	yes	yes	yes	yes
N	1053	1053	1053	1053	1053	1053
R^2	0.836	0.881	0.839	0.976	0.976	0.976

Note: Robust standard errors clustered at city in parentheses. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

5. Discussion

In this paper, through a theoretical framework and empirical analysis, we conclude that the construction of a CBEC pilot zone can promote economic growth. Through further analysis, we find that, on the one hand, the policy effect of a pilot zone is heterogeneous in different regions and different batches. On the other hand, CBEC can promote economic growth through three channels: urban digital construction, trade openness, and industrial agglomeration of the information service industry. Compared with the existing literature, our analyses of the policy effect of the establishment of pilot zones are more comprehensive.

From the results of the theoretical mechanism analysis, our paper finds that the “no ticket tax exemption” policy implemented by various pilot zones in terms of tax supervision helps facilitate the import and export activities of CBEC enterprises. This is consistent with the research findings of Martens [9] and Einav et al. [10], who found that breaking down barriers, such as those of consumption tax and value added tax, is conducive to the development of cross-border online trade. In addition, the simplification of the existing customs clearance process in each pilot zone is also conducive to the development of CBEC, which is consistent with the conclusion of Kim et al. [11].

When comparing our empirical results with the results of the existing literature [16,17], it must be pointed out that we arrived at a new conclusion regarding how the establishment of a CBEC pilot zone affects economic growth. We found that the construction of a comprehensive pilot zone enables economic growth through the three channels of urban digital construction, trade openness, and industrial agglomeration of the information service industry. We also deconstructed the policy effects of comprehensive pilot zones from the perspective of regional heterogeneity and different batches and described in detail the policy dividends generated by the establishment of comprehensive pilot zones. In the past, analysis of how the establishment of comprehensive pilot zones affected economic growth remained at the theoretical level, and the facts are still unknown. From an empirical point of view, our paper answers the key questions of how the establishment of comprehensive pilot zones will benefit economic development and which economic sectors will prosper.

More broadly, our causal framework quantitatively analyzes the economic effects of the construction of CBEC pilot zones. However, the social and environmental effects of the construction of CBEC pilot zones are still unclear; for example, the impact of the construction of a pilot zone on the employment demands of enterprises and on urban carbon emissions, which is a direction for further research.

6. Conclusions and Suggestions

6.1. Conclusions

Our paper regards the implementation of the CBEC comprehensive pilot zone policy as a "quasi-natural experiment". First, we theoretically analyzed the impact mechanism of the CBEC pilot zone policy on economic growth. Second, based on the data of Chinese cities from 2011 to 2019, we employed the DID model to evaluate the impact and internal mechanism of the establishment of CBEC pilot zones on economic growth. Third, we considered the heterogeneity of the impact of the CBEC pilot zone policy on economic growth in different locations and batches. The empirical results show the following:

(1) The construction of CBEC pilot zones has significantly promoted China's economic growth. To verify the effectiveness and robustness of the benchmark regression results, we used a parallel trend hypothesis test, a placebo test, and other methods.

(2) The impact of the construction of CBEC pilot zones on economic growth is significantly heterogeneous: first, compared with inland areas, the construction of CBEC pilot zones will play a clearer role in promoting economic growth in coastal areas; second, compared with the central and western regions, the construction of CBEC pilot zones will play a more evident role in promoting economic growth in the eastern region; and third, the economic effects of the construction of the first three batches of CBEC pilot zones are relatively apparent.

(3) The construction of CBEC pilot zones promotes economic growth through three channels: urban digital construction, trade openness, and the agglomeration of the information service industry.

6.2. Suggestions

Based on the above conclusions, we further propose the following policy recommendations:

(1) The construction of China's CBEC comprehensive pilot zones also requires a large amount of government policy support to promote the institutional innovation and model innovation of CBEC pilot zones. China's CBEC pilot zones have only been operating for a short time, and they still lack sufficient coping capacity in terms of operational stability and sustainability. Therefore, the Chinese government needs to design top-level policy support for the development of CBEC and provide support from the earliest aspects of CBEC development, with regard to things such as production land, commodity quality inspection, tax rate reduction, and fiscal and financial support. This will help CBEC pilot zones in the cultivation of large CBEC platform subjects, talent recruitment, system innovation, and model innovation.

(2) CBEC comprehensive pilot zones also need to continue to carry out bold reform and exploration. First, CBEC pilot zones should rely on modern information technology, blockchain termination, big data, and other digital technology to transform the service, logistics, and payment links in CBEC transactions and promote the innovation and development of CBEC business process re-engineering and supervision modes. Second, CBEC comprehensive pilot zones should continue to facilitate the integration of CBEC customs clearance, information sharing, and regulatory innovation to realize the liberalization and convenience of CBEC trade and to ultimately encourage CBEC to become a new growth point and a new competitive advantage for China's international trade development. Finally, CBEC pilot zones need to strive to build a complete

CBEC industrial chain and ecological chain and to form a CBEC online and offline business industry cluster.

(3) CBEC comprehensive pilot zones in different regions and batches need to identify their positioning and weaknesses to narrow the gap between different CBEC pilot areas. For example, Shenzhen has shortcomings in CBEC services. It is necessary to pay attention to the payment, logistics, and customs clearance systems of CBEC development, to optimize the service system, and to speed up the construction of CBEC industrial parks in Futian, Yantian, and other blocks. Shanghai and Hangzhou need to focus on how to further innovate the development and reform of CBEC based on the existing development advantages. Other CBEC pilot zones should pay heed to their resource endowment advantages; make full use of the CBEC pilot zone policy dividends in combination with regional characteristics; accelerate the construction of software and hardware, such as infrastructure, talent training, and policy supporting services for the development of CBEC; and accelerate the construction and development of CBEC pilot zones.

(4) The CBEC pilot zone is a good experimental field for institutional innovation. The Chinese government should actively refine the rules and standards to conform to the world digital trade development trend in the process of experimental exploration. On the one hand, this will continuously create convenient conditions for CBEC enterprises, reduce trade costs, improve trade efficiency, and accelerate the transformation and upgrading of China's CBEC. On the other hand, it will also help export the "Chinese-style template" of the CBEC development model to the world, along with the digital trade rules and standards, thus promoting global economic governance.

This paper still has some limitations. Due to data restrictions, we were not able to empirically study the economic effects of the comprehensive pilot zones established over the last two years. Furthermore, we were not able to study the economic effect of the construction of CBEC pilot zones from the micro perspective of enterprises.

Author Contributions: Conceptualization, M.Z. and Z.W.; Data curation, M.Z.; Formal analysis, M.Z. and Z.W.; Investigation, M.Z.; Methodology, M.Z.; Software, M.Z.; Writing—original draft, M.Z.; Writing—review and editing, Z.W. and X.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Research Project of the National Natural Science Foundation of China (grant number 71774129) and the Shaanxi Education Science Planning Project (grant number SGH21Y0007).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data in our paper are all from the *China Urban Statistical Yearbook* (https://data.cnki.net/yearbook/Single/N2022040095, accessed on 10 March 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Goldmanis, M.; Hortacsu, A.; Syverson, C.; Emre, O. E-Commerce and the Market Structure of Retail Industries. *Econ. J.* **2010**, *120*, 651–682. [CrossRef]
2. Lendle, A.; Vezina, P.-L. Internet Technology and the Extensive Margin of Trade: Evidence from EBay in Emerging Economies. *Rev. Dev. Econ.* **2015**, *19*, 375–386. [CrossRef]
3. Ma, S.; Guo, X.; Zhang, H. New Driving Force for China's Import Growth: Assessing the Role of Cross-Border e-Commerce. *World Econ.* **2021**, *44*, 3674–3706. [CrossRef]
4. Stoerk, T.; Dudek, D.J.; Yang, J. China's National Carbon Emissions Trading Scheme: Lessons from the Pilot Emission Trading Schemes, Academic Literature, and Known Policy Details. *Clim. Policy* **2019**, *19*, 472–486. [CrossRef]
5. Chan, K.S.; Xu, X.; Gao, Y. The China Growth Miracle: The Role of the Formal and the Informal Institutions. *World Econ.* **2015**, *38*, 63–90. [CrossRef]
6. North, D.C. *Institutions, Institutional Change, and Economic Performance*; Cambridge University Press: Cambridge, UK, 1990. [CrossRef]

7. Valarezo, Á.; Pérez-Amaral, T.; Garín-Muñoz, T.; Herguera García, I.; López, R. Drivers and Barriers to Cross-Border e-Commerce: Evidence from Spanish Individual Behavior. *Telecommun. Policy* **2018**, *42*, 464–473. [CrossRef]
8. Ho, S.-C.; Kauffman, R.J.; Liang, T.-P. A Growth Theory Perspective on B2C E-Commerce Growth in Europe: An Exploratory Study. *Electron. Commer. Res. Appl.* **2007**, *6*, 237–259. [CrossRef]
9. Martens, B. *What Does Economic Research Tell Us about Cross-Border E-Commerce in the EU Digital Single Market? Digital Economy Working Paper*; Institute for Prospective Technological Studies: Seville, Spain, 2013. [CrossRef]
10. Einav, L.; Knoepfle, D.; Levin, J.; Sundaresan, N. Sales Taxes and Internet Commerce. *Am. Econ. Rev.* **2014**, *104*, 1–26. [CrossRef]
11. Kim, T.Y.; Dekker, R.; Heij, C. Cross-Border Electronic Commerce: Distance Effects and Express Delivery in European Union Markets. *Int. J. Electron. Commer.* **2017**, *21*, 184–218. [CrossRef]
12. Wang, Y.; Wang, Y.; Lee, S.H. The Effect of Cross-Border E-Commerce on China's International Trade: An Empirical Study Based on Transaction Cost Analysis. *Sustainability* **2017**, *9*, 2028. [CrossRef]
13. Falk, M.; Hagsten, E. E-Commerce Trends and Impacts across Europe. *Int. J. Prod. Econ.* **2015**, *170*, 357–369. [CrossRef]
14. Lu, B.; Wang, H. Research on the Competitive Strategy of Cross-Border E-Commerce Comprehensive Pilot Area Based on the Spatial Competition. *Sci. Program.* **2016**, *2016*, 6216052. [CrossRef]
15. Jin, F.; Yan-Ling, C.; Xiao-Jun, J.; Liang, Z. An Observation on China Comprehensive Pilot Areas for Cross-Border E-Commerce in Henan Province. *Procedia Manuf.* **2019**, *30*, 77–82. [CrossRef]
16. Chen, N. Analysis of the Correlation between Cross-Border E-Commerce and Economic Growth Based on Hierarchical Multilevel Gray Evaluation Model. *J. Math.* **2022**, *2022*, 8455404. [CrossRef]
17. Wang, H.-D.; Zheng, C.-F.; Xiao, X. An AMOS Model for Examining the Factors Influencing the Development of China Cross-Border E-Commerce Comprehensive Pilot Areas. *Discret. Dyn. Nat. Soc.* **2022**, *2022*, e1889589. [CrossRef]
18. Han, Y. Impact of Environmental Regulation Policy on Environmental Regulation Level: A Quasi-Natural Experiment Based on Carbon Emission Trading Pilot. *Environ. Sci. Pollut. Res.* **2020**, *27*, 23602–23615. [CrossRef] [PubMed]
19. Han, F.; Xie, R.; Lu, Y.; Fang, J.; Liu, Y. The Effects of Urban Agglomeration Economies on Carbon Emissions: Evidence from Chinese Cities. *J. Clean. Prod.* **2018**, *172*, 1096–1110. [CrossRef]
20. Billmeier, A.; Nannicini, T. Assessing Economic Liberalization Episodes: A Synthetic Control Approach. *Rev. Econ. Stat.* **2013**, *95*, 983–1001. [CrossRef]
21. Zang, J.; Wan, L.; Li, Z.; Wang, C.; Wang, S. Does Emission Trading Scheme Have Spillover Effect on Industrial Structure Upgrading? Evidence from the EU Based on a PSM-DID Approach. *Environ. Sci. Pollut. Res.* **2020**, *27*, 12345–12357. [CrossRef]
22. Laitsou, E.; Kargas, A.; Varoutas, D. The Impact of ICT on Economic Growth of Greece and EU-28 under Economic Crisis. In Proceedings of the 2017 Internet of Things Business Models, Users, and Networks, Copenhagen, Denmark, 23–24 November 2017; IEEE: Copenhagen, Denmark, 2017; pp. 1–6.
23. Edquist, H.; Henrekson, M. Swedish Lessons: How Important Are ICT and R&D to Economic Growth? *Struct. Chang. Econ. Dyn.* **2017**, *42*, 1–12. [CrossRef]
24. Liu, Q.; Qiu, L.D. Intermediate Input Imports and Innovations: Evidence from Chinese Firms' Patent Filings. *J. Int. Econ.* **2016**, *103*, 166–183. [CrossRef]
25. Pierce, J.R.; Schott, P.K. The Surprisingly Swift Decline of US Manufacturing Employment. *Am. Econ. Rev.* **2016**, *106*, 1632–1662. [CrossRef]
26. Bertrand, M.; Duflo, E.; Mullainathan, S. How Much Should We Trust Differences-in-Differences Estimates? *Q. J. Econ.* **2004**, *119*, 249–275. [CrossRef]
27. Baron, R.M.; Kenny, D.A. The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *J. Personal. Soc. Psychol.* **1986**, *51*, 1173. [CrossRef]
28. O'Donoghue, D.; Gleave, B. A Note on Methods for Measuring Industnial Agglomeration. *Reg. Stud.* **2004**, *38*, 419–427. [CrossRef]
29. Freedman, M.L. Job Hopping, Earnings Dynamics, and Industrial Agglomeration in the Software Publishing Industry. *J. Urban Econ.* **2008**, *64*, 590–600. [CrossRef]
30. Beck, T.; Levine, R.; Levkov, A. Big Bad Banks? The Winners and Losers from Bank Deregulation in the United States. *J. Financ.* **2010**, *65*, 1637–1667. [CrossRef]
31. La Ferrara, E.; Chong, A.; Duryea, S. Soap Operas and Fertility: Evidence from Brazil. *Am. Econ. J. Appl. Econ.* **2012**, *4*, 1–31. [CrossRef]
32. Abadie, A.; Diamond, A.; Hainmueller, J. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *J. Am. Stat. Assoc.* **2010**, *105*, 493–505. [CrossRef]
33. Liu, Y.; Luan, L.; Wu, W.; Zhang, Z.; Hsu, Y. Can Digital Financial Inclusion Promote China's Economic Growth? *Int. Rev. Financ. Anal.* **2021**, *78*, 101889. [CrossRef]
34. Solomon, E.M.; van Klyton, A. The Impact of Digital Technology Usage on Economic Growth in Africa. *Util. Policy* **2020**, *67*, 101104. [CrossRef] [PubMed]

sustainability

Article

Augmenting Community Engagement in City 4.0: Considerations for Digital Agency in Urban Public Space

Michael G. Hunter [1,*], Alessandro Soro [1], Ross A. Brown [1], Joel Harman [1] and Tan Yigitcanlar [2]

[1] School of Computer Science, Faculty of Science, Queensland University of Technology, 2 George Street, Brisbane, QLD 4000, Australia
[2] School of Architecture and Built Environment, Faculty of Engineering, Queensland University of Technology, 2 George Street, Brisbane, QLD 4000, Australia
* Correspondence: mg.hunter@qut.edu.au

Abstract: An engaged community that reflects a diverse set of experiences is key to an equitable and livable city. However, maximizing engagement activities is often difficult when competing with residents' busy schedules and hectic daily lives. To explore new opportunities in this space, we developed four augmented reality experiences to learn more about the potential for this technology to transform community engagement practices in the context of City 4.0. City 4.0 utilizes digital technologies to transform public services and the local economy. Its goal is to produce more sustainable urban and societal outcomes. Our findings suggest that augmented reality is least successful when used to recreate existing engagement practices, such as surveys or questionnaires, and more successful when it empowers a sense of agency and ownership over the process in its users. The way augmented reality situates information can aid in making public space feel personal to the individual. In this way, augmented reality's affordances are less about overlaying digital information in physical space and more about how this can enable individuals to reclaim a sense of control and relevance in the relationship between citizens and councils. We aim to contribute: (a) novel interaction paradigms and an evaluation of their effectiveness and limitation, and (b) new insights into how to support citizens' sense of agency in public discourse with augmented reality. This paper highlights the value of augmented reality's affordances to bring to light new interactions between community engagement stakeholders.

Keywords: augmented reality; digital agency; urban public space; community engagement; placemaking; human-computer interaction; smart city; City 4.0; urban planning and design; Brisbane

Citation: Hunter, M.G.; Soro, A.; Brown, R.A.; Harman, J.; Yigitcanlar, T. Augmenting Community Engagement in City 4.0: Considerations for Digital Agency in Urban Public Space. *Sustainability* **2022**, *14*, 9803. https://doi.org/10.3390/su14169803

Academic Editor: Nikos A. Salingaros

Received: 19 May 2022
Accepted: 6 August 2022
Published: 9 August 2022

1. Introduction

The last decade has presented many challenges for cities around the world. First and foremost, urban migration continued to rise as projections estimated the world's urban population to be as high as 68%. As of 2020, North America's urban population is estimated at 83%, Australia at 80%, and Europe at 75% [1]. The past two years have also seen the COVID-19 pandemic challenge city resilience and slow forward-looking policy in response to handling the immediate emergency and uncertainty about a future 'post-COVID-19' recovery [2]. Rapid population growth in cities often requires changes in the urban landscape to accommodate for new housing, mobility, and other shared services. These changes can often affect large amounts of the pre-existing population. For these changes to be well-suited to both new and existing populations, city administrations will often look to community engagement strategies that hope to inform and consult with those most affected. In doing so, city administrations can attempt to ensure their decision is more defensible against critiques and the project is less likely to encounter roadblocks or pushback. The concept of City 4.0 introduces a new paradigm in which city administrations utilize digital technologies to connect all city stakeholders in a way that produces more sustainable urban outcomes. Furthermore, community engagement is a broad term used in an urban context to describe the different types of engagement from citizens in urban issues [3]. While research can often discuss this in terms of participatory planning [4–6], we utilize the term community engagement because it is a term shared by both researchers and city administrations alike.

Research on community engagement puts a strong focus on the politics of participation and how different approaches to community engagement can have vastly different results [7,8]. Participatory and collaborative approaches particularly look at broadening the base of stakeholders involved, and especially at 'giving a voice' to traditionally marginalized groups [9,10]. The community engagement literature has seen a few interesting new developments in participatory design methodologies [11], interventions within public space [9,12], or applications of novel technologies [4] that address the challenge of inclusion. Furthermore, the COVID-19 pandemic has accelerated the adoption of QR codes and remote communication, whilst Facebook's announcement of their 'metaverse' plan brings concepts of mixed reality and blockchain to the mainstream, offering new opportunities for technologically mediated interactions in community engagement processes. Likewise, the new and popular city blueprints, such as smart cities or City 4.0, also underline the crucial importance of community engagement [13–16]. City 4.0 utilizes technological developments and digitalization to transform local public services and the local economy. It leverages these digital technologies and data to connect citizens, producing sustainable and desired urban, environmental, and societal outcomes for all.

In the context above, augmented reality (AR) is gaining attention as an enabler of situated engagement [17], improved urban conversation [18], and participation in cities [4]. AR allows digital information to be situated in physical space, so in an urban context, development details can be shared between stakeholders at the physical location or displayed and embodied at scale. The main affordance of AR—to overlay digital information over the physical world—is touted as a key driver in encouraging engagement with planning proposals, or in consulting with citizens about future developments [5,18,19]. Whilst there has been research into what this can mean for co-design in participatory planning, it is noted that the greater political systems in place still ultimately decide the face of participation in any given engagement process [4]. As such, a critique of AR or the implementation of any novel technology in cities is that, while the technology may offer new interactions, without a fundamental shift in the political relationship between councils and citizens, the technology will simply offer new opportunities for top down 'engagement theatre' and consensus building activities.

Our research studied four different AR experiences intended to explore bottom-up and middle-out forms of engagement. The experiences ranged from the recreation of existing community engagement methods in AR to entirely new interaction paradigms used to blend physical and digital.

- The city builder: This experience presents a list of options that allow the user to build their own city in augmented reality. Each option queues a separate musical loop, so that they end with a unique city and song. This experience was intended to test how AR could offer new interactions for gathering feedback from the community and visualize the results of citizen choices back to the participant in real time.
- The city spaces quiz: This experience acted like an augmented reality photo gallery, showing photos of space designed for cars or for people. It was designed to highlight the large amount of space required by cars, compared to people. Users would be quizzed on what level of scale they thought the photo represented. We intended this to be a form of interactive and educational tool that would help us gauge how receptive individuals are to education delivered through this medium.
- Bridges for Experience: This experience displayed 3D models of different bridges overlayed across a map of Brisbane River. Showcasing the potential to demonstrate future development plans to citizens in AR. Like the above experience, we wanted to

understand how users would react to an AR experience that mimicked more traditional community consultation strategies.

- The portals: This final experience displayed a portal in physical space that users could walk through to enter a virtual city. They could walk around this virtual city and experience a virtual world situated in the same space as the physical world.

These experiences were offered as 'probes' to demonstrate different interaction paradigms in context and understand the opportunities and challenges of adopting AR technology in community engagement practices. We offered these experiences firstly to participants as part of a city-wide STEAM festival that lasted three weeks and secondly in a half-day workshop with four participants.

What emerged from our findings is that a key value of augmented reality for improving community engagement practices is not just about the unique interaction paradigms it affords, but more so about what opportunities it represents for the individual citizen within the community engagement political dynamic. Participants often discussed the feelings of agency that our city builder and portal experiences offered, and that while the interaction itself was novel, it was more that the participant felt that their voice was heard and that they were contributing. Participants were much less interested in our one-way informative experiences like the gallery, and much more interested in the experiences that allowed them to create for themselves.

Therefore, we contribute that, while AR affords us interesting opportunities to overlay digital information in physical locations, this is perhaps only relevant in community engagement when it empowers a sense of agency in the individual engaging with it. Simply put, a recreation of existing engagement processes in AR are unlikely to be successful purely because of AR's affordances, however, AR's unique affordances do allow for an improved sense of agency for individuals in public space and a rethinking of the greater engagement experience. This was evident in the responses of our participants between the city builder and the Bridges for Experience. While the Bridges for Experience and city builder both use AR to overlay digital information on the physical world, the city builder was much more positively received due to the information being created and customized to hold significance to the end user.

Below, we explore the literature surrounding community engagement, participatory planning, and urban human–computer interaction (HCI). We do this to explain our use of the term community engagement and situate this focus within the participatory planning literature. Furthermore, it is important for us to draw on the urban HCI literature to contextualize our findings and show how urban HCI research has previously conceptualized the use of AR. Following this, we discuss our early conversations with stakeholders, our reasoning for and design of each study and the results that we found. We then highlight the areas of interest that emerged from these studies before finally discussing some interesting directions for future urban AR research.

2. Literature Background

2.1. Community Engagement and Participatory Planning

We start this related works section by highlighting the literature surrounding community engagement and participatory planning. In most urban research, participatory planning is the term used to highlight the relationship between councils and citizens in relation to the development of urban areas [6]. However, the term participatory can carry slightly different meanings that shift the focus of research in the area. In the context of participatory design, participatory planning will often focus on the individual and design interventions, such as media façades, urban screens, and mixed reality, that encourage participation in design with individuals from the bottom up [9,20]. In other cases, the focus of participatory planning research is more political, analyzing varying levels of civic participation in relation to greater democracy. In particular, Legacy's [8] paper highlights the way a large majority of the participatory planning literature tends to analyze top-down and bottom-up perspectives and how they can affect participation. While these topics certainly are not mutually exclusive, it is worth noting how the focus can shift from participation as an individual design activity to participation as a civic duty or participation within political process.

Smith [6] presents an interesting summary of this, defining the rational and consensual aspects of participatory planning. The rational aspect considers that "individuals are more intimately involved with environmental changes; they can provide a planning process with information and judgements regarding local systems". The consensual aspects consider "societal units, being involved in the determination of planning processes related to the domain of that societal unit, which may lead to a further integration of power with authority, a move toward a democratic society". This theoretical basis for participatory planning [6] posits these two aspects in a way that has been accurately reflected in the analysis of research for years to come. We relate the rational aspects more closely to participatory design research, and the consensual aspect as the political focus of participation.

One constant in participatory planning research, however, is its focus on planning processes. That is, most of the participatory planning research understandably conceptualizes participatory planning as something related to a specific project that will invite change in the urban area. A gap that arises in this literature is that its strong focus on planning processes and the politics of participation narrow its focus in such a way that it can miss the other motivations for engagement exhibited by citizens. To further this, community engagement literature has covered engagements between different stakeholders in medical research, such as the cultural barriers regarding the uptake of a vaccine [21], or in education research, to help embed cultural knowledge in an educational curriculum [22]. Aligning with the community engagement terminology, we are able to draw insights from city administration practice and the broader engagement literature perspectives, inclusive of participation, planning, and politics, but not restricted to these lenses. With this in mind, we utilize the term community engagement for two reasons: (a) community engagement is not specifically tied to the planning and development of an urban area, but more so to citizens' engagement with councils, and (b) community engagement is often utilized as a tool by city administrations within participatory planning processes. In the first instance, community engagement allows us to investigate the relationship between councils and citizens from a broader perspective, rather than in relation specifically to urban planning. Secondly, often in practice, community engagement is the terminology used for strategies that encourage broader participation. In this paper, we often refer to traditional community engagement strategies and therefore feel it relevant to use this terminology as our basis for comparison.

2.2. Top Down, Bottom Up, and Middle Out

When analyzing community engagement strategies and the relationship between city administrations and citizens, there are often three different relationships that are discussed: top down, bottom up, and more recently, middle out.

Top-down relationships position engagement as led by city administrations or governments, and often focuses on the way that city administrations consult with communities or deliver information to communities [8,23]. This form of engagement, when critiqued, is said to be more performative or see the role of the citizen as tokenistic [24,25]. This is often because the decisions regarding the planning have been made, and community engagement is used as a strategy to inform citizens of the decisions. In other cases, it is found that the policy environments and power dynamics between varying levels of government can often sideline community objectives [26].

Bottom-up relationships position engagement as something that empowers individuals to create, design, and actively participate in interventions at a grassroots level [23,27]. While it could be said that some bottom-up interventions could still be empowered by city administrations, they are typically led by a community group, social movement, or individuals and look to collaborate on decision making according to the chosen intervention. In this way, community engagement is a more collaborative process than the post-decision-making process of the top-down perspective. However, sometimes overlooked in these positions is the challenge of scale and the individual's perceived relevance of the project [26].

Bottom-up projects tend to be driven by individuals or community groups, and the shared purpose of that group adds perceived relevance of the intervention to all members of that group [3]. The projects of city administrations, however, can sometimes be so large that they affect a much wider group of the population and, as a result, it is increasingly difficult to ensure the engagement feels personal and relevant to everyone affected [3]. The final body of literature follows middle-out engagement.

Middle-out engagement looks to draw on the collective knowledge of all actors to provide opportunities for collaborative community engagement processes. The pop-up interventions of Fredericks and Caldwell [9] enlisted the help of both councils and local community stakeholders in their design, implementation, and deployment. These interventions utilized the strengths of both groups of stakeholders to ensure the interventions could be deployed at scale and for the benefit of broader councils, whilst still drawing on the knowledge and design of individuals at a local level to ensure their relevance and value to that local community. More local, state, and national urban policy is moving in a similar direction with a recent white paper from England's Ministry of Housing, Communities and Local Government proposing that better information be delivered to local communities, and technologically mediated solutions be developed that allow for a more democratic system between residents, communities, entrepreneurs, businesses, and councils [28]. Furthermore, research by Usavagovitwong et al. [29] highlights the concept of 'community architects' across Asia, specifically demonstrating the value of architects in enabling a link between poor communities, local organizations, planning and development agencies, and broader government initiatives. In both of these works, the value that comes from enabling engagement between all city stakeholders and adopting a middle-out engagement approach is made clear.

Our research aimed to explore the value of a middle-out engagement approach, by partnering with local councils to host our digital experiences and offering interactions that specifically elicited feedback and knowledge from individuals and included their participation through creation within the experience. We ultimately wished to explore how this approach can develop into more conversational platforms between citizens and councils, where the middle-out ethos can be coupled with urban HCI interventions that contribute to a broader city platform. We summarize the three approaches in Table 1 below.

Table 1. Engagement approaches.

Approach	Definition	Advantages	Disadvantages
Top-Down	Led by city-administration. Tends to deliver information regarding planning decisions.	Can deliver information at scale, and utilize existing IT infrastructure.	Often feels 'tokenistic' as citizens are not included in decision making. Little engagement from citizens as perceived as irrelevant and impersonal.
Bottom-Up	Led by individuals or community groups, designs grass-roots solutions with citizens.	Relevance to particular group, further engagement due to personal feel. Collaboration before decision making.	Difficult to scale, solutions specific to smaller urban groups. Often niche issues, and under-resourced.
Middle-Out	Aims to utilize knowledge of all actors, enlisting the help of councils to facilitate, and individuals to contribute.	Utilizes the value of all stakeholders, facilitates a relationship between stakeholders that is usually challenged.	Limited previous work to draw on. Broader group of stakeholders makes project planning and execution much more difficult.

2.3. Urban Human–Computer Interaction

Urban HCI is often discussed in research in both the context of community engagement and the context of social movements and digital activism. Like the way in which participation can be viewed through a political lens and a more design lens, the intersection of public space and technology often explores the way technology can shape the political landscape at a grassroots level, and the way individuals can use technology to create and design their own communities or experiences within that space [30]. Vadiati [31] highlighted the shift technology has on public space into an augmented urban space continuing beyond its physical boundaries. Furthermore, the way this matter affects urban governance was discussed, noting the narrative across research that ICTs are ultimately activating more citizens who would not engage in urban matters through traditional outlets [31].

While the influence of communicative technologies in urban space is increasingly evident, there are still many facets to explore around their implementation, the interactions they offer, and how they may or may not shift the current relationship between citizens and city administrations. Analysis of digital activism and citizen science [18,32–34] has highlighted the power for communicative technologies to empower individuals and social movements, such as the global effects of #MeToo and #BlackLivesMatter; however, the fact that these issues can transcend national boundaries can sometimes be to the detriment of their relevance or impact at the local level [18]. Alternatively, e-participation research will often investigate digitally mediated participation at a more local or state level, although this tends to revert to a focus on participation in planning or governance processes [4,35,36]. Ultimately, while there is a growing amount of research looking into the implementation of novel technologies at a local level, most of the focus lands on how these technologies can augment planning processes.

AR has been utilized many times in recent years as a tool to test new co-design methods for city planning. Its strengths as a visual communicator—allowing users to place objects and visualize proposals—is often touted as a key reason for its value in co-design methods, and its ability to run on modern mobile devices is seen as an incentive for younger audiences. In Bandung, Indonesia, augmented reality was used as a learning tool for future environmental planning. In this way, augmented reality allowed for more interactive storytelling that could combine local folklore with environmental challenges and was found to be adopted by the students in such a way that they could communicate to other community groups through augmented reality to collaborate on solutions to environmental challenges and educate those less aware of environmental issues [37].

Furthermore, in New Zealand, Allen et al. [38], developed an application that allowed members of the public to visualize 3D models of new building designs at their proposed physical location. The participants responded extremely positively to using AR as a visual tool in this way. Since 2011, many similar studies that use AR have taken place to co-design urban spaces, and in particular buildings and future developments [5,39]. Lastly, a paper by Saßmannshausen et al. [4] highlighted the value of these AR tools as extensions of community engagement practices that can entice a younger audience’s participation. While this work was still grounded in planning activities, it also explored how AR can be used as an informative tool, a co-determination, and a co-design tool. In this context, it was not just about visualizing existing plans, but about encouraging participation in the design of these tools that would then visualize information.

One aspect that we find particularly interesting is how using AR for participatory planning can open up to new co-design possibilities outside the immediate realm of planning. That is, by enabling users to place and visualize digital objects in physical space, we can also enable new possibilities for collaboration between these individuals in public space. Furthermore, the development of algorithmic techniques allow for a procedural generation of building designs, so that architectural expertise can be generated without the need for expertise from the individual citizen [40]. In particular, Potts [41] analyzed the way PokemonGo and augmented reality games (ARGs) could activate public spaces, increasing community interaction and facilitating the exploration of a city. Furthermore, numerous studies have explored the impact of AR to reappropriating public space [42,43]. These studies highlight the way these tools can be used for empowering individuals in urban space, not just in a planning context, but in the broader relationship between citizens, city administrations, and public space. Our research seems to sit at the intersection of these few topics.

In sum, we use the term community engagement because we understand AR's affordances to extend beyond participation in planning activities and empower more broad community engagement. Having said that, we are not purely focused on the political empowerment of communities in opposition to city administrations in the city–citizen political paradigm. Lastly, while we recognize the value of AR as a visual tool in cities, in this work, we wish to explore what that means for improving the engagement of an individual within their community and public space, especially considering the smart city or City 4.0 context.

3. Methodology and Analysis

The research took place over a period of about 12 months across several meetings with relevant stakeholders and two main studies. Our team consisted of two interaction design researchers and a post-doctoral game developer with expert knowledge of Unity3D and ARCore. This research is part of a broader exploration of augmented sociality [44] that seeks to find new opportunities for a community-oriented, user-generated mixed reality. Augmented reality itself is most easily understood as technology that overlays digital information over the physical world [18].

Preliminary research into smart cities [45] highlighted a shift in focus for city administrations from implementing technologies to address the assumed needs of urban challenges to instead using communicative technologies to talk to citizens and understand what challenges existed [18].

Given that our aim was to explore new opportunities for a socialized AR, we reached out to local stakeholders to understand their perspective on the state of current community engagement practices. For sake of clarity, we will summarize the main insights gained from the stakeholder meetings in this section, because these informed the design of the four AR probes adopted in the subsequent studies. The findings from the two main studies are presented and discussed in the next sections.

The first study was conducted in the wild [46,47] during a public festival. As Rogers noted, research in the wild "is likely to reveal more the kinds of problems and behaviors people will have and adopt if they were to use a novel device at home, at work, or elsewhere" [46]. While studies in naturalistic settings often follow preliminary lab studies, we decided to conduct this preliminary exploration in a relatively uncontrolled setting with the aim to reach a wide public audience and gain an understanding of general expectations, technical challenges, and public interest.

The second study, conducted over six months after the public exhibition, consisted of a half-day workshop with four participants, during which specific thoughts on community engagement and the AR probes were shared and discussed. The workshop adopted an approach inspired by future technology workshops [48,49] and cross-cultural dialogical probes [50].

3.1. Stakeholders Consultations

Initial discussions involved employees of the Brisbane city council in various offices with responsibilities spanning across community engagement and digital services. Since these consultations involved all participants in their professional roles, with a view of discussing possible collaborations and partnerships, these conversations were not conducted under the project's ethical framework and no 'informed consent' declarations were collected at the meetings. We nevertheless omit reporting on their detailed roles or positions to maintain confidentiality. No audio or video recordings were made of these meetings and the summary below is based on the authors' detailed notes taken during the meetings. To ensure the participants' viewpoints are correctly represented, we shared a draft of this paper with them, seeking comments and inviting corrections.

The discussions focused on the technologies that were in use or that the stakeholders were potentially interested in and how these played a role in different community engagement strategies. We also discussed, in their view, how augmented reality may be utilized to address the challenge of community engagement at scale.

3.1.1. Opportunities

It emerged that collaboration and engagement were among the biggest use cases for introducing technological innovation in public projects, with a particular focus on connected community spaces and socialized or virtual platforms. The principle of fostering agency and sense of ownership was an important driver, in line with existing initiatives currently supported through more traditional means.

One important aspect was the drive to support existing community groups in implementing local engagement programs and to remain financially viable, rather than exclusively focus on large scale centralized support. Supporting these localized groups involves a deep understanding of how they operate, both internally and within the complex ecosystem of municipal offices and regulations.

Overall, a specific interest emerged for tools and applications to help deliver infrastructure projects. This involves supporting various goals and phases in a typical infrastructure project, from seeking feedback in the designs, informing the public on intended outcomes and benefits, and engaging with the community to help understand the project and create the best possible experience. This is within the understanding that councils ultimately own the assets and high-level planning decisions have typically already been made at the time consultation begins. It was noted that these processes are currently mostly supported by rather low-tech tools, often paper-based, or at best, online services.

3.1.2. Challenges

It also emerged that the downside of innovation rests in its inherent high level of risk. Decisions on adopting technological innovation are very sensitive to political cycles, and the availability of resources and funds can be ephemeral.

A particular emphasis was placed on the journey to deploy specific technologies within established council procedures. The need for infrastructure, the challenge of data management, and the community expectations and understanding of new technologies all pose problems.

A challenging goal also emerged from drawing in those who are hesitant and support the accessibility on a community scale with the aim of making the city more inclusive to vulnerable and homeless people. Symmetrically, an important but challenging goal was identified to explore new ways to gather, analyze, and use data from community engagement tools and initiatives.

3.2. Augmented Reality Probes

Following our preliminary stakeholder meetings, we distilled several design inspirations that we further developed through discussion in the research team and based on the relevant literature and previous works:

- Engage new audiences and members of the public who are otherwise hesitant to participate in community initiatives;
- Explore synergies with data mining, visualization, and data driven decision making;
- Explore unique affordances of augmented reality, especially the appeal of visual interactive tools, localized contents and interactions, and similarities or resonance with virtual reality;
- Maintain a focus on deployment and accessibility by a broad audience.

While developing the AR probes, an opportunity arose to present them to a very large audience as part of the Curiocity festival. This STEAM festival has a strong focus on using science, technology, engineering, arts, and math to curate interesting public exhibits throughout Brisbane. Its goal is to encourage people to navigate around the city from experience-to-experience learning about both the city and the various applications of

STEAM. The festival ran for three weeks throughout March and hosted over 60 installations, at least two thirds of which were utilizing augmented and virtual reality.

Curiocity aligned extremely well with our initial research agenda as it allowed us to showcase the experiences to the public and to gauge an interest in the public for augmented reality usage under the broader theme of the festival.

However, this opened our research to a whole new spectrum of challenges. Of course, being hosted by the Curiocity festival, the location of our installation was decided for us, which was outdoor. Secondly, being a part of a broader festival meant our participants would not be solely engaging with our exhibition and therefore would have little prior understanding of our installation and its purpose. Challenges like these shaped a few design decisions as it was important that we developed something intuitive, playful and that could work outdoors.

A range of ideas were brainstormed and assessed based on likely environmental challenges, availability of material or software libraries, COVID-19 constraints for sharing devices or manipulating surfaces. We considered offering a city builder experience, a gallery or quiz experience, a table with a Lego-city and sensors that trigger music, and a final AR wall that would overlay digital information on top of a Brisbane city map. These can be seen in Figure 1.

An important factor in refining the design ideas was to offer a selection of different AR interactions and replicate typical engagement methods. The quiz would be an interesting way to explore how councils could inform citizens in AR, whilst the map would show models in AR and highlight the potential for AR as a visualization tool. However, the city builder, which had been chosen as the first experience, was instead intended to transform previous engagement processes and highlight the way interactive games could elicit the same information as a survey in a more playful way.

Printing Markers

For these AR experiences to work on location, we had to decide whether we would use image recognition or GPS. We could either use spatial anchors to place the experiences in a physical space so that they would be triggered when a user pointed their phone at that location, or we could use image recognition and image markers so that the experience would trigger when the user pointed a device at a particular image. We decided on image recognition for the three table experiences, the city builder, gallery, and model bridges, and spatial anchors for the portal. A spatial anchor was more useful for the portal as it required the user to walk through it, so we could not have a physical obstacle in the way of that interaction.

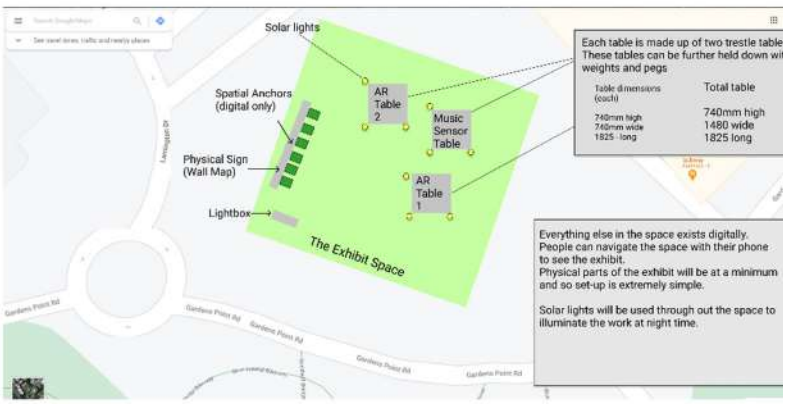


Figure 1. *Cont.*

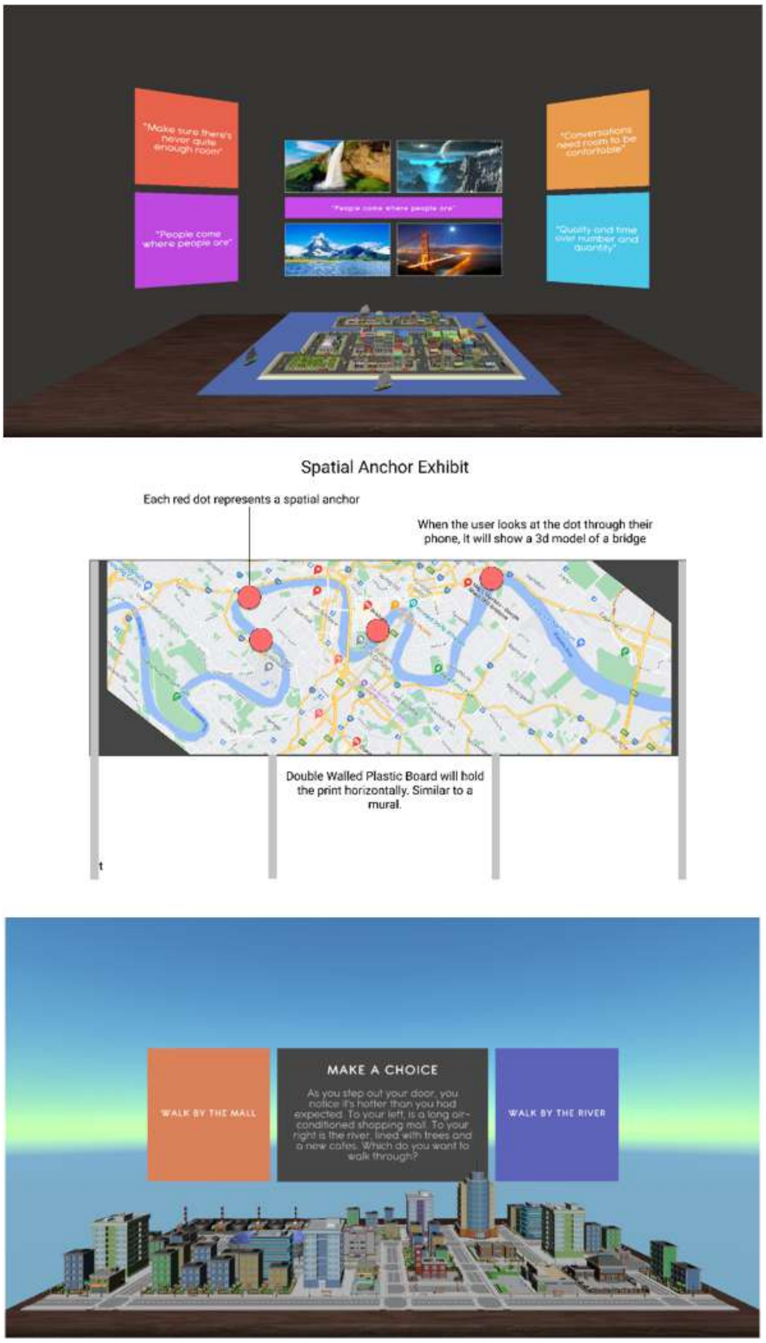


Figure 1. Initial designs brainstormed for the Curiocity Festival.

We wanted to design the markers in such a way that they were recognizable by the image recognition software, whilst still interesting for the participants and indicative of the experience on offer. We decided to double the markers as instructions and measured them out so that they would cover the entire tables we had hired. The markers can be seen in Figure 2.

These markers had to be printed on a specific type of plastic board so that they were water resistant and would not warp in the outdoor area under different weather conditions. The weather resistance of the board was useful, although this added a reflective layer to the marker that caused issues with the image recognition once placed outside. Again, not being able to test these on location, the challenges of image recognition did not become apparent until the markers were in place.

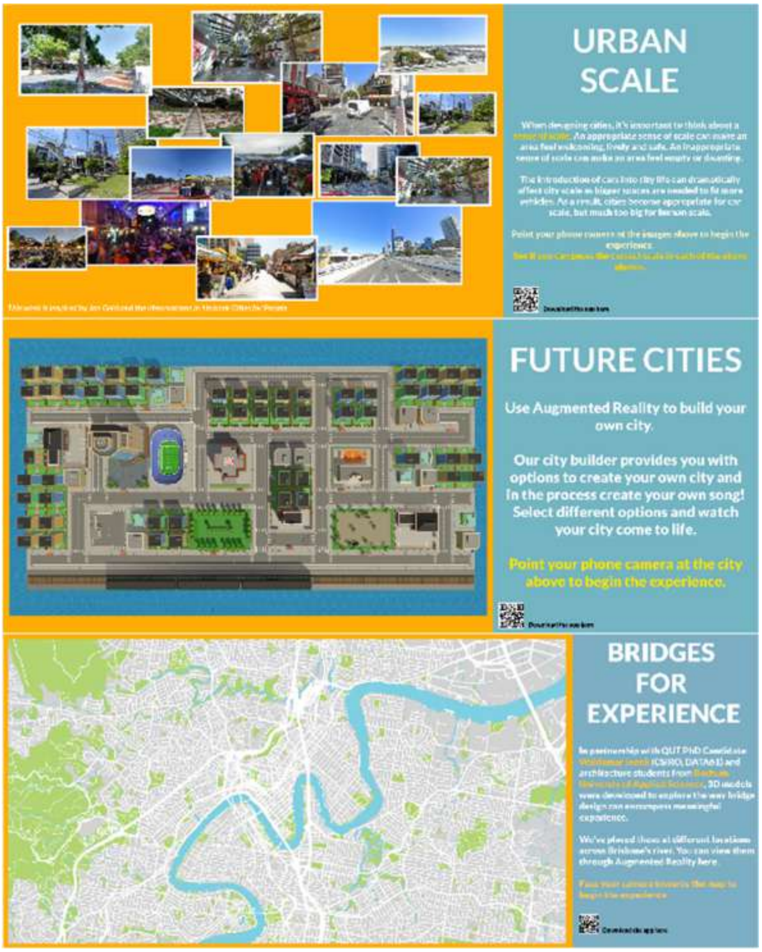


Figure 2. Image recognition designs brainstormed for the Curiocity Festival.

3.3. The Experiences

'Future Cities for Future People' is an Android smart phone application that can be downloaded from the Google Play Store to play. The name is an intended reference to Jan Gehl [51], who largely inspired the ideas presented throughout the AR experiences. The experiences within the application are as follows.

3.3.1. City Builder

The first game is a city builder. We present an empty city grid to the user and ask them to make choices about what should be in the city. Each choice presented is also linked to a musical loop. Once the participant makes all of their choices, they are presented with a unique city and an accompanying song. The choices that we presented to the user are as shown in Table 2.

The choices that we presented to the participant are of course not mutually exclusive in a real-world city planning context and, furthermore, we did not present them with any bias as to what may or may not be the correct answer. In some cases, we could visualize the results of choices, such as parking lots and public transport, wherein the parking lots choice would take up much more city space than if public transport was selected. However, it was not our intention to attempt to simplify what may be extremely complex issues. Instead, it was about presenting a questionnaire in an interesting and playful way to see if there was potential for this kind of conversation between city administrations and citizens moving forward. With the user's consent, their interaction could be recorded so that the choices could be processed and analyzed. Again, this was intended to demonstrate how these interactions could spur real-time conversation between councils and citizens.

Table 2. City builder choices.

Lay Roads	
Standard housing	Highrise buildings
Public services	
Parking lots Supermarket Stadium Opera house Stadium	Public transport Market stalls Observation tower Tourist hotel Gardens

Once users had built their city, they could also activate a street view by clicking on a section of road. This would transport the user to that part of the road, and rather than look at the grid, they were free to move their phone around in space and view the city they had built as if they were standing in it. This can be seen below in Figure 3.

Figure 3. The city builder screenshots of third person and first person street view.

3.3.2. Bridges for Experience

The second experience displays a map of Brisbane and its river, pictured in Figure 4. Across the river are interactable buttons that the user can press to see different 3D model bridges. These bridges were designed by students from Bochum's University of Applied Sciences as part of an assessment exploring the education of students within media architecture and their approach to bridge design. The experience allows the bridges to be viewed in 3D space hovering above the map of Brisbane and displays a small description as to why the bridges were designed as they were. This experience is intended to showcase the possibilities for community consultation through AR. Using AR, we could quickly demonstrate future bridge designs to the community, hoping to engage them and demonstrate a way in which councils could engage with communities regarding future urban planning designs. Furthermore, it showcases the possibilities for university, industry, and council partnerships as 3D models can be exported from architecture programs and displayed within AR applications.

Figure 4. Bridges for Experience.

3.3.3. City Spaces

The city spaces quiz—shown in Figure 5—is intended to be an interactive game that would help citizens think differently about a sense of scale in urban space. Taking much inspiration from Jan Gehl's book Cities for People [51], this game acts like a photo gallery in 3D space, displaying different photos of Brisbane's city environments to the participant. It showcases urban areas that are designed for cars, humans, and both and aims to educate participants on the difference between human needs and needs for cars and the amount of space required for each.

3.3.4. City Portals

The final experience is a GPS located AR portal. This portal exists at the physical location of the installation and can be interacted with by simply walking through it. Users can hold their camera up to the space to see the portal appear, and then walk through to experience an entirely virtual world. The photos in Figure 6 demonstrate how once inside the portal, the users' field of view becomes enveloped in the virtual environment. Here we hoped to demonstrate the affordances of AR as a situated and embodied interaction method.

3.4. The Participants

3.4.1. Study 1: Curiocity

Over the course of the 17-day festival, there were 90 downloads of the application. During the festival there was constant rain for roughly 5 days, and a COVID-19 lockdown for another three. As this installation was outside, this certainly played a part in the number of interactions we received. Of the 90 downloads, we received consent 125 times to receive data about the participants choices. This number is greater because users were prompted every time they opened the application, so in some cases the participant may have used the application more than once, and therefore consented each time they used it. Of the 125 data logs retrieved, only 43 contained data regarding the users' specific interactions with the application. The remaining 82 display a 'TEST' input, so that we know the application was opened but no further data was recorded. This tells us that no further experiences were triggered, which could be through user choice, or through a break in the applications functionality, such as the AR image recognition failing to work correctly.

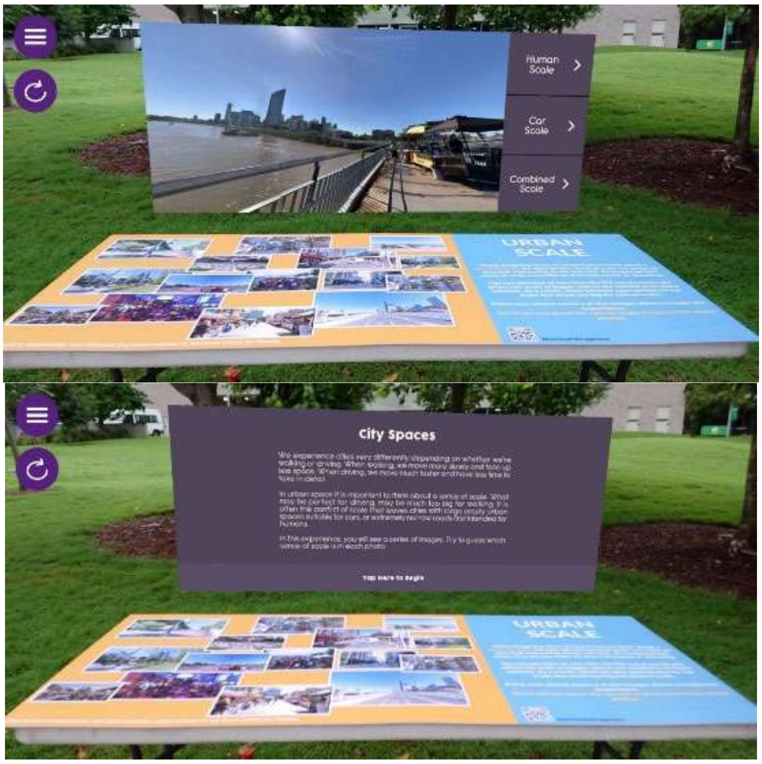


Figure 5. The city spaces quiz in augmented reality above the physical table markers.

3.4.2. Study 2: Workshop

This workshop took place indoors at our university campus with 4 participants, 2 men and 2 women, recruited via email. The participants were students, aged between 20 and 35, studying either computer science, architecture, or design. Each received a gift voucher as a token of our appreciation for their participation. Informed consent and image release forms were signed by all participants, and video and audio was recorded by the research team.

With all the challenges that the first study presented, it became clear that a second study was required to help gain further qualitative insights and allow for testing of the application in an indoor controlled setting. It was our aim to learn more about how participants interacted with the experiences at a qualitative level in a setting that would allow for a more focused discussion around their attitudes towards community engagement and augmented reality and allow them to test our application with support from our team in case there were any errors. The nature of Study 1 raised a few challenges and issues that Study 2 was able to compensate for. This workshop was inspired methodologically by Future Technology Workshops by Vavoula et al. [49]. In this particular methodology, the participants envisage future technologies and work backwards to co-design through current state activities. In our case, we wanted to break the pre-existing expectations surrounding augmented reality and community engagement to help them think about what community engagement could look like with this technology, rather than just what these technologies could replicate. The workshop ran in three sessions as outlined below:

- The first session ran for roughly 50 min and was strongly focused on how individuals perceived community engagement, and whether their perception of community engagement was always political. The discussion aimed to highlight the types of activities that the participants would engage with in their day-to-day life to understand what they might constitute as valid forms of participation just as much as political forms of participation.
- The second session ran for roughly 25 min and allowed each of the participants to use a tablet or phone to play with our AR experiences inside the workshop room. We had prepared each of the experiences on a separate table, and otherwise provided little instruction for the participants. We were particularly curious to see their entire interaction with the applications, from whether they would be able to understand and use them to their feedback and experience with them.
- After this, in the third session, we reconvened for a discussion surrounding the application to understand the participants experiences with the AR experiences, their attitudes towards them, and their thoughts on how or if they could be applied in a future community engagement context.

Figure 6. The city portal.

3.5. Analysis

Our research approach utilized both quantitative and qualitative methods. Our application can record all the choices that the user makes within. With the users' consent, we were able to record how they interact with each experience, the choices they make within that experience, and their total time spent in each experience. This provides us with an interesting quantitative picture, which was particularly useful in our first study when interviews were limited.

In the second study, we recorded the audio of each of the discussions as well as the video and photography of the participants' usage of our application. We attempted to use Dragon software to later transcribe the audio; however, the audio files failed to be recognized, so we transcribed the discussions ourselves. Of these discussions, we first read through the transcriptions separately, identifying some of our own perceptions of the data. We later had meetings to discuss our thoughts between authors as to the types of comments that had emerged from the transcriptions. Following this, we entered a more formal coding phase, reviewing the entire transcriptions multiple times to identify potential emerging themes.

4. Results and Findings

4.1. Curiocity Festival

Of the data collected, we recorded 206 total city builder choices and users entered a part of the city at the street view level 277 times. The average length of time users spent playing with the city builder was roughly 3 min. The remaining choices can be found below in Table 3.

Table 3. City builder choices.

Choice	Options	Number of Times Chosen	Most Popular
Choice 1	Standard housing	8	Highrise buildings
	Highrise buildings	15	
Choice 2	Parking lots	3	Public transport
	Public transport	19	
Choice 3	Supermarkets	5	Market stalls
	Market stalls	13	
Choice 4	Observation tower	15	Observation tower
	Stadium	3	
Choice 4	Opera house	9	Equal opera house and tourist hotel
	Tourist hotel	9	
Choice 5	Stadium	4	Gardens
	Gardens	13	

In this table, we have not included the button interactions that did not require the user to make a choice, such as the 'Lay roads', 'Public services' and 'Explore' interactions. These interactions were mandatory and required the user to press them to move onto the next part of the game.

The bridge experience was triggered 27 times and the choices were as shown in Table 4.

The city spaces quiz was triggered 29 times and received 74 separate attempts to respond to its questions. The quiz would normally end after six different questions, but sometimes users would stop early or play the quiz multiple times. We displayed photos of Brisbane at different scales to the participants and asked them to guess whether this part of the city was at car scale (a highway), human scale (a quad or public space), or a combined scale (a market stall that had overtaken a street). We then recorded their answers for each. Of the 74 separate responses, we had 34 incorrect responses and 40 correct responses.

Table 4. City builder choices.

Choice	Number of Times Selected
Bridge A	27
Bridge B	21
Bridge C	20
Bridge D	14
Bridge E	20

The final portal experience was triggered 40 times. This experience did not require any further interactions and therefore did not track any more information about the participant within the virtual environment. We simply recorded when they entered into and subsequently left the portal. The average time spent inside the virtual environment was roughly three minutes.

4.1.1. Observational Analysis

Firstly, many participants were not willing to download the application. Participants would walk past the installation or read about it, but upon realizing they needed to download an application, they would stop their engagement. In other cases, the augmented reality aspect of the application would carry perceived complexity and act as a deterrent for individuals. Two comments heard were, 'it is like the sims, except that was simple, now it's too technologically advanced', and 'maybe we will download it when we get home'. On average, roughly four to five people would walk past the exhibit each hour, but most would continue on to the next exhibit without downloading the application. Lastly, those who would interact with the installation and were happy to talk more about it were, interestingly, academics, architects, or retired couples.

Something that is perhaps not often discussed in the AR literature is the difficulty of delivering a consistent experience with the current technology, and even more so when it is setup outside and open to the public. In this scenario, the promise of using novel technology as part of a public festival to engage citizens completely blindsided us to the challenges of its successful implementation.

4.1.2. Technical Limitations

There were a few technical limitations that came with the development of this application.

Firstly, Android users had to have at least Android 8.0 or higher as the version of operating system on their device. A surprising number of devices were not compatible with this requirement, as this version of Android was only released in 2017.

Secondly, the quality of the camera on the device made a large difference as to whether it would recognize the markers in the physical space or not. Setting up this installation outdoors meant that the markers would catch a lot of sunlight. The markers were also printed on a PVC foam board so that in case of rain or intense weather, they would not warp, fade, or have the image and colors run. This board was useful in durability, but surprisingly reflective and so, in full sun, it would often reflect light back at the device's cameras. On modern smart phones this presented little issue, as the cameras on these devices are quite high quality. However, we used tablets at the exhibit to lend to users who were not interested in downloading the application, and the cameras on these tablets struggled immensely to recognize the markers in full sunlight and, as a result, users were often unable to properly start the experience.

Thirdly, application size, processing power, and overall performance were all challenges that had to be prioritized in developing an easily accessible experience. We were acutely aware that asking participants to download an application to their device would add a layer of difficulty to the experience that would result in some users not taking interest and, as a result, we aimed to keep the application size as small as possible. This meant lowering the polygon count and resolutions of many of the objects within the experience. We managed to reduce the download size to 87 mb for the entire application and by reducing some resolution sizes and polygon count, we also hoped to increase performance across varying devices. Even still, performance was widely varied between modern phones and modern tablets. Smart phones would generally run the experiences with no framerate drops, although the phones would increase in temperature considerably within just a few minutes of playing. The tablets would often experience framerate drops instantly with the city builder experience and again would overheat considerably. Sometimes this overheating would result in even more framerate drops, which would result in the camera being unable to register the marker as well and the experience would jitter in and out of frame.

Lastly, whilst not an immediate issue, the low polygon counts of the application meant that the objects in the experience were reasonably simple and restricted our choices for the city builder or bridges for experience to low polygon assets. These low-polygon assets in some ways enhanced the playful nature of the experience but may have been less conducive to a believable or meaningful tool for community engagement. It was difficult to convey a level of weight to the users' choices and, as a result, may not have educated the user on urban planning as much as presented a simple fun experience.

4.1.3. Environmental Challenges

The festival began on 12 March 2021 and ran through till 28 March 2021. Brisbane is a sub-tropical city that tends to experience 25-degree weather for the majority of the year. March 2021 ended up being a surprisingly extreme weather period for Brisbane and, during the first two weeks of the festival, there was almost non-stop rain. This greatly limited the number of participants and really tested the durability of the installation. Daily checks were required to ensure the markers had not moved, bowed, or collected too much water. When the rain ceased, the temperatures would reach upwards of 30+ degrees and during the middle of the day, the installation was so hot that it became almost completely unfeasible to expect participants to engage with our experiences. Furthermore, the reflectiveness of the markers would often confuse the image recognition software so that the augmented reality failed to trigger or the level of glare on the screen made it near impossible to see. While the installation was open to the public for three weeks, it only became usable for a few hours a day, a few days of the week. The final challenge was COVID-19, wherein during the final week of the festival, a spike in cases swept through Brisbane, eventuating in a snap-lockdown.

4.2. Workshop Discussions

Below, we present the data from our workshops. While the data below cover a few different topics, there is an overarching discussion around agency both in an emotional sense—how the participant feels in the relationship between city and citizen—and a physical sense—how the interactions and controls provide the participant with agency in that moment.

The idea of agency was often discussed throughout the three sessions, where participants stressed the importance of feeling control, feeling heard, and feeling like their opinions were valued. Throughout the first session, there were many positive comments made about the way one council member had turned budgetary ideas for the suburb into a survey that residents could complete.

Participant A: 'I remember a petition that he put out, which was very interesting for me, he sent it out to everybody who lives in that particular ward, and at that time I was living there, and it was a budget for improving different places within west end, and we as a resident could vote, and you didn't have to be a citizen, you just had to be a resident in west end, to vote where you wanted the money to go.

I thought that was very, very interesting because it was giving me agency in places that I use temporarily, but might not use in the future, and I still got a say right now'.

Furthermore, in the second and third sessions, participants gave a lot of positive feedback towards the city builder experience purely because of the agency it offered.

Participant B: "that's why I would say the build your own city is interesting because I feel like I had a say. Right? So, I do respond by building, creating something, that means that I'm active in the sense of like having a real ownership of what the city is going to look like. So, you do have that agency over, at least you have this feeling, that you have agency, whereas, when you think about just responding or being informed, I'm not too sure to what extent you are actually engaged, or do you participate".

Participant C: 'But I did like the opportunity to simulate my different choices, like, one city looks like this, and this is the music I get, and if I cancel, what happens if I do the opposite choices and it was very different and I found it rewarding to see my options matter in the game'.

Alternatively, discussions around the design of the interface were focused on how the AR controls removed agency in a way that made the participants feel confused, or unable to see the value in AR specifically.

Participant D: 'I just don't see why you need to make it AR for that experience, because I feel like you don't have much control, I mean from taking the perspective of a gamer, you have more control in the environment that you're in, but for that one, you can't even see to read the small text. Whereas, if it wasn't in AR, you could zoom, and you have more freedom'.

Participant C: 'Why do you need to walk around? When can you just sit down and move your mouse to navigate around the environment?'

Participant A: 'When it goes to the interface, I think the written things shouldn't be with the visual, they should pop up before so that I know exactly what I'm seeing, and then I go into that interface and see it, and it should just be there as backup because it's pretty difficult to read, because it's so small'.

The above comment was in relation to our design choice to include the interface in the augmented reality experience and that the instructions were floating positioned in physical space, rather than before or after the experience as a normal digital screen.

Participant C: 'I was just thinking, I would see it as you can build your city on a desktop or normal mobile interface, and then when you want to see the city, you can portal yourself there and I think that's the only place where the AR makes sense'.

A clear challenge in using novel technologies for community engagement is designing it in such a way that the user feels they have agency even though the technology is new and potentially complex. There are many factors in the design of the interface that can have unintended consequences, regarding the attitude of the user, that can shape their attitude towards the city administration overall. In our case, placing the interface within the AR resulted in some cases where participants felt that it was intentionally designed that way to give the illusion of agency without giving agency.

In this way, the interface and interactions of digital technologies are mediums through which to convey the attitudes or relationships from one stakeholder to another.

5. Discussion

The overarching aim of this research was to understand the potential for augmented reality to improve community engagement. The way augmented reality can allow users to visualize and interact with virtual environments in physical space creates a few interesting ways for councils to present information and modify consultation methods to feel more engaging. However, perhaps more important is the way that augmented reality can afford users agency over the space they occupy. Hence, the focus should not be on augmented reality solely as a technology, but instead as a new interaction paradigm that allows citizens an embodied interaction in the context of specific place. As the related works show, there have been several studies based on visualizing future plans through AR. Our findings suggest that the way AR can situate digital information in a physical space is not just interesting for its visualization affordances, but more so for how it can empower citizens within a public space.

The popularity amongst participants for our portal and city builder experiences, supported by the discussions during our final workshop session, highlights a clear desire for agency in community engagement processes. This element of community engagement is perhaps one of the more discussed aspects of civic participation processes, where critiques of 'engagement theater' and 'degrees of tokenism' often arise [8,25,52,53]. In the recent technology-enabled participatory planning literature, it is often explored how augmented reality can be used as a tool for more immersive visualizations and information sharing regarding planning projects [4,5]. It is hoped that a result of more immersive information and visualization is improved participation, especially from younger demographics [4]. A shortcoming that is often realized in this research, however, is that ultimately the issues with engagement do not rest on the technology, but instead stem from the institutional processes. Whilst AR may offer more meaningful forms of visualization, it does not necessarily offer any more meaningful engagement because in some cases the planning processes do not actually allow for citizen input.

There is a large amount of research demonstrating the success of AR visualizations for artists and improved culture within a city [42,43]. This was further highlighted by the success of several AR art visualizations offered during the festival that our first study participated in. However, we suggest that these successes are not purely due to the novelty of overlaying digital information on the real world, but instead due to what it represents in the relationship between individuals, councils, and public space. Digital urbanism and placemaking research highlight the use of AR in 'guerilla' settings for individuals to protest within or regain ownership of a public space [54,55]. In these contexts, AR is often positively discussed due to the way it empowers individuals to place their own art or visualizations over the physical world, whereas, in top-down community engagement processes, technology is often discussed as potentially novel, but ultimately still limited by a few challenges presented by the political nature of the process or the perception of the public towards the city administration that implements it.

We find this to be especially true if AR is used to recreate existing forms of community consultation. During both our time at Curiocity festival and our workshop discussions, our least popular AR interactions were those that visualized existing digital information in physical space but did not offer any further customization or creation. Those that delivered information in one-way from the application to the users struggled to convey relevance to the participants or generate excitement from the users even with the novelties of AR. We see this as a natural extension to the sentiment in the paper by Saßmannshausen et al., where it seems that a lot of the successes found were regarding the co-design and collaboration involved with developing the AR prototypes [4]. While the AR technology helped enable interaction with a younger audience, the participatory design methodologies that were employed appear to be a key success factor in the audience responding positively to the applications [4].

5.1. Most Engaging Interaction Paradigms

We suggest that the value of AR regarding community engagement requires a re-conceptualization of its affordances. While the novelty of overlaying digital information in physical space may be useful for appealing to younger audiences [4], our findings appear to show that this holds most significance when it affords users the chance to create or have some control over the information they are experiencing. In this case, the visual and interaction affordances typical to AR seem to be most effective when they are used to afford a deeper level of agency in the user. When designing interactions for AR within community engagement, we suggest that a strong focus should first be on delivering a sense of agency or control. In the context of public space, AR can allow users to overlay digital information over a public space, allowing them agency over a public space in a way previously only possible through street-art or perhaps graffiti. This was reflected in our interviews with participants:

> "that's why I would say the build your own city is interesting because I feel like I had a say. Right? So, I do respond by building, creating something, that means that I'm active in the sense of like having a real ownership of what the city is going to look like"

A criticism often received in this context is 'why AR?'. Most users do not feel that simply overlaying digital information on the physical world really adds extra value. Instead, these interactions become valuable when they allow the user to create or act in a way that makes the digital information relevant to them as users. Out of the four interactions that we explored in AR, we found that the experiences that recreated existing council–citizen interaction paradigms were overwhelmingly the least popular. The city spaces quiz and model bridge gallery—which both used AR to visualize information—struggled to engage participants in the same way the city builder and portal did. Participants commented on the way that these experiences felt particularly 'one-way.' In these experiences, participants were able to use AR as a novel way to overlay digital information on the real world, but because they did not have any input into the creation of this information, they struggled to see its relevance. While other research suggests that AR can be used to extend existing community engagement processes [4], our findings suggest that perhaps an extension alone is not enough if these processes do not initially afford a sense of control to the end user.

The experiences that were the most popular were those that afforded the most agency. Participants thoroughly enjoyed the city builder and portal experiences and favored less the city spaces quiz or model bridge experience. Interestingly, whilst these experiences were less relevant to the broader city in that they were not representative of future plans within Brisbane and were not visualizing photos of Brisbane's public spaces, they were much more relevant to the individual who had created them. Whilst AR certainly enabled them to visualize their city in physical space, the key information is that it was 'their' city. The level of agency over their creation continued to be a resounding positive comment.

5.2. Augmented Reality for Citizen Centered Public Space

When conceptualizing AR within processes of community engagement, it is key to consider AR not just as a technology for visualizing council-defined information, but as a way for citizens to reclaim elements of public space and a sense of agency within the broader planning discussions. This is not to negate the clear affordances of AR in situating and overlaying digital information in physical spaces to create more engaging experiences, but instead to recognize that successful citizen engagement with these tools is also closely linked to the way they allow citizens to have control over a part of the discussion and appropriate the AR tools to promote their own projects and pursue their own agenda.

By placing more emphasis on the alternative ways that augmented reality can enable agency within urban planning, then our design choices are not necessarily linked to visualization or information sharing and can begin to open up to all sorts of unique, meaningful and playful interactions. When our participants entered into the street view of the city they had created, they were more interested in sharing with each other the types of buildings that they had chosen. These low-poly buildings were not necessarily exciting to look at, but the excitement came from the conversations that were enabled because of the choices made by the participant. In this experience, AR enabled a situated platform for conversation, more so than a space for visualizing existing council plans.

Participant A: 'we're using AR and then discussing about it, you have one medium it's a conversation starter, you sit and discuss'.

Our initial research [18] aimed to explore the types of conversational platforms that AR could afford. Initially, we suggested that AR offers three affordances: (a) It could be used as a situated real-time enabler of conversation between councils and citizens; (b) it allows for visual, physical, embodied, and practical co-design, and; (c) it enables a data-driven reflection of the above processes. Our city builder concept aimed to demonstrate these three affordances in that it allowed participants to co-design their ideal city, not only through text-based interactions, but also as visual digital objects in a physical space. Using choices, we attempted to emulate the types of conversations that could be had between these two stakeholders and, lastly, by recording the data of consenting users, we were afforded a reflective quantitative picture of the conversations that had taken place. While an unintended finding was the deeper desire for agency that users discussed, we do believe that the city builder experience successfully explored what an AR platform for conversation could look like and the types of interactions that could maximize engagement from individuals. While its focus was on city building, its interaction methods demonstrated the ways that AR's visual affordances could be used to share information between stakeholders, not just about planning, but about broader city environmental or social concepts.

5.3. Challenges and Limitations

Once this deeper sense of agency is established, however, it is then difficult not to undermine it with underexplored interaction methods. Minor errors in interaction design can bleed into the users feeling of agency in such a way that their attitudes towards the entire engagement process and the technology are then tarnished. Whilst in some cases, our participants thoroughly enjoyed the way they could move throughout physical space while inside a virtual environment, they were equally frustrated by the choices we made with interface and controls and struggled to see the value in AR as a standalone technology. It is a challenge for future research to explore how AR platforms can create a sense of agency in the citizen–council relationship and, at the same time, ensure that the methods of interaction do not undermine this feeling.

Interestingly, participants often suggested that a sole focus on AR was too strong, and that a more hybrid approach to AR should be utilized instead, using existing technologies for many aspects of the engagement process and only using AR for a particular use case where its strengths were obvious. Where we had kept our different interaction methods segmented, instead, the participants were more interested in a combination of the experiences that saw them building a city, using a portal to explore it, and using the gallery to explore the impacts of their choices.

5.4. Future Research Directions

In this way, we see AR as an extension of the number of smart-city conversational platforms that have been developing around the world. The smart city literature has noted the success of developing platforms for conversation first before large-scale implementations of technology [18,45,56]. Cities such as Amsterdam have demonstrated much greater success on a considerably smaller financial budget by developing platforms that allowed citizens, institutions, and councils to engage with each other and decide upon projects that were most important for the city [45,57]. For now, these platforms tend to exist mostly online, and once projects have been decided, then move towards physical labs and other consultation spaces.

We identified a gap in the literature for discussing AR's position in a broader set of city technology platform infrastructure. Common uses of AR focus so strongly on its ability to transform previous planning or educational activities that it is not often considered how AR can be integrated into the broad set of technologies that are already implemented in cities. As sensors, online websites, social media, and artificial intelligence are all widely researched in their implementation and interoperability for improving city maintenance, resource efficiency, or civic participation, it should be further researched where augmented reality can sit in this greater tech-stack. AR on its own may sometimes struggle to convey usefulness to individuals outside of its novelty, but perhaps, in using city-wide sensor data and artificial intelligence, AR could be used to present real-time data to individuals to enable more informed conversations.

Participant B: 'the city builder, the portal, the quiz, like the merging of these three is very useful in the sense that, and this is, I'm thinking from the perspective of, educating, making decisions, having some agency and then council getting something out of it'.

Much research has been undertaken in the past five years that investigates the idea of City as a Platform. This research often investigates the way that open data, communicative technologies, and a multistakeholder participatory approach can be used to solve new environmental, social, and economic urban challenges [48–50]. However, at this stage, the platforms and research have a strong focus on artificial intelligence and the sharing of open data and little focus on the value of augmented reality to situate or visualize this information. When AR is discussed, it is more utilized in small scale, standalone, local prototypes. An interesting avenue for future research is to understand how AR can be used as a part of the broader city platform. Where the concept of open data may only be useful to those with the knowledge to analyze and use it, perhaps AR could present this data in new, meaningful ways that encourage participation from non-technologists. Furthermore, the situated and embodied affordances of AR represented in our portal experience highlight the way that these platforms could shift from something digital and found only online, to something experienced in a physical space and interacted with by individuals in a more embodied way. For those perhaps too busy to interact with websites, surveys, and digital town halls, AR could augment the existing spaces that citizens inhabit in their day-to-day practices, shifting engagement from something required of individual citizens to something that is instead observed by city administrations or individuals within these spaces.

6. Conclusions

Augmented reality (AR) is often considered most valuable in an urban context when it can be used to visualize future plans for the area, especially in the context of a smart city or City 4.0. The successes of most research in this area, however, tend to highlight how much enjoyment and engagement the participants have in creating and placing their own models and buildings. Our findings suggest that while the visualization novelties of AR are exciting for end users, what is most engaging is the sense of agency or control that can be found from the creation and ownership of digital content in a physical space. City 4.0 suggests that more sustainable outcomes can evolve from a digitalized connection of multiple city stakeholders. Furthermore, numerous case studies [4,20,28,29,32] demonstrate that by maximizing citizen engagement in urban processes, more sustainable solutions are generated (such as improving representation of poor communities so that they can plan their own urban futures relative to their unique social and cultural processes). When we consider AR's affordances in the relationship between citizens and city administrations, rather than just the affordances it can offer visually, we can start to understand the powerful role of AR in an urban context and the value it can bring to community engagement processes when designed correctly. Its ability to enable agency and contextualize information in an embodied and visual way can further empower citizens and improve council data and, as a result, create more equitable and sustainable outcomes for urban environments. Finally, we suggest further research is required to understand AR's role within the broader urban technological infrastructure, acting as a medium for conversation between citizens and city administrations, rather than just a visualization tool for small scale urban change.

Author Contributions: Conceptualization, M.G.H. and A.S.; methodology, M.G.H.; software, J.H.; validation, A.S., R.A.B. and J.H.; formal analysis, M.G.H.; investigation, M.G.H.; resources, A.S.; data curation, M.G.H.; writing—original draft preparation, M.G.H.; writing—review and editing, A.S., R.A.B. and T.Y.; supervision, A.S. and R.A.B.; project administration, A.S.; funding acquisition, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Australian Research Council, grant number DP200102478.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank the editor and anonymous referees for their constructive comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. How Has the World's Urban Population Changed? | World Economic Forum. Available online: https://www.weforum.org/agenda/2020/11/global-continent-urban-population-urbanisation-percent/ (accessed on 1 February 2022).
2. Hunter, M. Resilience, Fragility, and Robustness: Cities and COVID-19. *Urban Gov.* 2021; *in press*. [CrossRef]
3. Balestrini, M.; Rogers, Y.; Hassan, C.; Creus, J.; King, M.; Marshall, P. A City in Common: A Framework to Orchestrate Large-Scale Citizen Engagement around Urban Issues. In Proceedings of the Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; Association for Computing Machinery: New York, NY, USA, 2017; Volume 2017, pp. 2282–2294.
4. Saßmannshausen, S.M.; Radtke, J.; Bohn, N.; Hussein, H.; Randall, D.; Pipek, V. Citizen-Centered Design in Urban Planning: How Augmented Reality Can Be Used in Citizen Participation Processes. In Proceedings of the DIS 2021: ACM Designing Interactive Systems Conference: Nowhere and Everywhere, Virtua, 28 June–2 July 2021; Association for Computing Machinery, Inc.: New York, NY, USA, 2021; pp. 250–265.
5. Bhardwaj, P.; Joseph, C. V Ikigailand: Gamified Urban Planning Experiences For Improved Participatory Planning. A Gamified Experience as a Tool for Town Planning. In Proceedings of the IndiaHCI 2020: IndiaHCI'20: 11th Indian Conference on Human-Computer Interaction, Online, 5–8 November 2020. [CrossRef]

6. Smith, R.W. A Theoretical Basis for Participatory Planning. *Policy Sci.* **1973**, *4*, 275–295. [CrossRef]
7. Shaw, M. Stuck in the Middle? Community Development, Community Engagement and the Dangerous Business of Learning for Democracy. *Community Dev. J.* **2011**, *46*, ii128–ii146. [CrossRef]
8. Legacy, C. Is There a Crisis of Participatory Planning? *Plan. Theory* **2017**, *16*, 425–442. [CrossRef]
9. Fredericks, J.; Caldwell, G.; Tomitsch, M. Middle-Out Design: Collaborative Community Engagement in Urban HCI. In Proceedings of the 28th Australian Conference on Computer-Human Interaction—OzCHI'16, Launceston, Australia, 29 November–2 December 2016; ACM Press: New York, NY, USA, 2016.
10. Blanco, C.; Kobayashi, H. Urban transformation in slum districts through public space generation and cable transportation at northeastern area: Medellin, colombia. *J. Int. Soc. Res.* **2009**, *2*, 84.
11. Klerks, G.; Hansen, N.B.; O'Neill, D.; Schouten, B. Designing Community Technology Initiatives: A Literature Review. In Proceedings of the OzCHI'20: 32nd Australian Conference on Human-Computer Interaction, Sydney, Australia, 2–4 December 2020; pp. 99–111. [CrossRef]
12. Schroeter, R.; Foth, M.; Satchell, C. People, Content, Location: Sweet Spotting Urban Screens for Situated Engagement. In Proceedings of the Designing Interactive Systems Conference, DIS'12, Newcastle upon Tyne, UK, 11–15 June 2012; ACM Press: New York, NY, USA, 2012; pp. 146–155.
13. Yigitcanlar, T.; Cugurullo, F. The Sustainability of Artificial Intelligence: An Urbanistic Viewpoint from the Lens of Smart and Sustainable Cities. *Sustainability* **2020**, *12*, 8548. [CrossRef]
14. Malek, J.A.; Lim, S.B.; Yigitcanlar, T. Social Inclusion Indicators for Building Citizen-Centric Smart Cities: A Systematic Literature Review. *Sustainability* **2021**, *13*, 376. [CrossRef]
15. Rotta, M.J.R.; Sell, D.; dos Santos Pacheco, R.C.; Yigitcanlar, T. Digital Commons and Citizen Coproduction in Smart Cities: Assessment of Brazilian Municipal e-Government Platforms. *Energies* **2019**, *12*, 2813. [CrossRef]
16. D'Amico, G.; L'Abbate, P.; Liao, W.; Yigitcanlar, T.; Ioppolo, G. Understanding Sensor Cities: Insights from Technology Giant Company Driven Smart Urbanism Practices. *Sensors* **2020**, *20*, 4391. [CrossRef]
17. Fredericks, J.; Fredericks, J. From Smart City to Smart Engagement: Exploring Digital and Physical Interactions for Playful City-Making. In *Making Smart Cities More Playable*; Springer: Berlin/Heidelberg, Germany, 2019.
18. Hunter, M.; Soro, A.; Brown, R. Enhancing Urban Conversation for Smarter Cities—Augmented Reality as an Enabler of Digital Civic Participation. *Interact. Des. Archit.* **2021**, *48*, 75–99. [CrossRef]
19. Shen, J.; Wu, Y.; Liu, H. Urban Planning Using Augmented Reality. *J. Urban Plan. Dev.* **2001**, *127*, 118–125. [CrossRef]
20. Foth, M. Participatory Urban Informatics: Towards Citizen-Ability. *Smart Sustain. Built Environ.* **2018**, *7*, 4–19. [CrossRef]
21. Carnegie, E.; Whittaker, A.; Brunton, C.G.; Hogg, R.; Kennedy, C.; Hilton, S.; Harding, S.; Pollock, K.G.; Pow, J. Development of a Cross-Cultural HPV Community Engagement Model within Scotland. *Health Educ. J.* **2017**, *76*, 398–410. [CrossRef] [PubMed]
22. Riley, L.; Howard-Wagner, D.; Mooney, J.; Kutay, C. Embedding Aboriginal Cultural Knowledge in Curriculum at University Level through Aboriginal Community Engagement. *Divers. High. Educ.* **2013**, *14*, 251–276. [CrossRef]
23. Breuer, J.; Walravens, N.; Ballon, P. Beyond Defining the Smart City. Meeting Top-down and Bottom-up Approaches in the Middle. *TeMA-J. Land Use Mobil. Environ.* **2014**, *8*, 155–156. [CrossRef]
24. Arnstein, S.R. A Ladder of Citizen Participation. *J. Am. Plan. Assoc.* **1969**, *35*, 216–224. [CrossRef]
25. Bowen, F.; Newenham-Kahindi, A.; Herremans, I. When Suits Meet Roots: The Antecedents and Consequences of Community Engagement Strategy. *J. Bus. Ethics* **2010**, *95*, 297–318. [CrossRef]
26. Sanga, N.; Gonzalez Benson, O.; Josyula, L. Top-down Processes Derail Bottom-up Objectives: A Study in Community Engagement and 'Slum-Free City Planning'. *Community Dev. J.* **2021**, *37*, 9–10. [CrossRef]
27. Niederer, S.; Priester, R. Smart Citizens: Exploring the Tools of the Urban Bottom-Up Movement. *Comput. Support. Coop. Work. CSCW Int. J.* **2016**, *25*, 137–152. [CrossRef]
28. Ministry of Housing Community and Local Government. *Planning for the Future*; Ministry of Housing Community and Local Government: London, UK, 2020.
29. Usavagovitwong, N.; Luansang, C. Housing by People: Performance of Asian Community Architects. 2010. 81p. Available online: https://www.academia.edu/5103694/Housing_by_People_Performance_of_Asian_Community_Architects (accessed on 18 May 2022).
30. Pancholi, S.; Yigitcanlar, T.; Guaralda, M. Public Space Design of Knowledge and Innovation Spaces: Learnings from Kelvin Grove Urban Village, Brisbane. *J. Open Innov. Technol. Mark. Complex.* **2015**, *1*, 13. [CrossRef]
31. Vadiati, N. Alternatives to Smart Cities: A Call for Consideration of Grassroots Digital Urbanism. *Digit. Geogr. Soc.* **2022**, *3*, 100030. [CrossRef]
32. Brown, A.; Franken, P.; Bonner, S.; Dolezal, N.; Moross, J. Safecast: Successful Citizen-Science for Radiation Measurement and Communication after Fukushima. *J. Radiol. Prot.* **2016**, *36*, S82–S101. [CrossRef] [PubMed]
33. Halupka, M. Clicktivism: A Systematic Heuristic. *Policy Internet* **2014**, *6*, 115–132. [CrossRef]
34. George, J.J.; Leidner, D.E. From Clicktivism to Hacktivism: Understanding Digital Activism. *Inf. Organ.* **2019**, *29*, 100249. [CrossRef]
35. Lim, S.B.; Yigitcanlar, T. Participatory Governance of Smart Cities: Insights from e-Participation of Putrajaya and Petaling Jaya, Malaysia. *Smart Cities* **2022**, *5*, 71–89. [CrossRef]

36. Purwandari, B.; Hermawan Sutoyo, M.A.; Mishbah, M.; Dzulfikar, M.F. Gamification in E-Govemment: A Systematic Literature Review. In Proceedings of the 2019 4th International Conference on Informatics and Computing, ICIC 2019, Semarang, Indonesia, 16–17 October 2019; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2019.
37. Argo, T.A.; Prabonno, S.; Singgi, P. Youth Participation in Urban Environmental Planning through Augmented Reality Learning: The Case of Bandung City, Indonesia. *Procedia-Soc. Behav. Sci.* **2016**, *227*, 808–814. [CrossRef]
38. Allen, M.; Regenbrecht, H.; Abbott, M. Smart-Phone Augmented Reality for Public Participation in Urban Planning. In Proceedings of the OzCHI'11: 23rd Australian Computer-Human Interaction Conference, Canberra, Australia, 28 November–2 December 2011; ISBN 9781450310901.
39. Imottesjo, H.; Thuvander, L.; Billger, M.; Wallberg, P.; Bodell, G.; Kain, J.H.; Nielsen, S.A. Iterative Prototyping of Urban Cobuilder: Tracking Methods and User Interface of an Outdoor Mobile Augmented Reality Tool for Co-Designing. *Multimodal Technol. Interact.* **2020**, *4*, 26. [CrossRef]
40. Postle, B. Brunopostle/Urb/Wiki/Homemaker—Bitbucket. Available online: https://bitbucket.org/brunopostle/urb/wiki/Homemaker (accessed on 18 May 2022).
41. Potts, R.; Jacka, L.; Yee, L.H. Can We 'Catch 'Em All'? An Exploration of the Nexus between Augmented Reality Games, Urban Planning and Urban Design. *J. Urban Des.* **2017**, *22*, 866–880. [CrossRef]
42. Biermann, B.C.; Seiler, J.; Nunes, C. The AR | AD Takeover: Augmented Reality and the Reappropriation of Public Space. 2011. Available online: Publicadcampaign.com (accessed on 9 February 2022).
43. Liao, T.; Humphreys, L. Layar-Ed Places: Using Mobile Augmented Reality to Tactically Reengage, Reproduce, and Reappropriate Public Space. *New Media Soc.* **2015**, *17*, 1418–1435. [CrossRef]
44. Soro, A.; Brown, R.; Wyeth, P.; Turkay, S. Towards a Smart and Socialised Augmented Reality. In Proceedings of the Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–8.
45. Desouza, K.C.; Hunter, M.; Jacob, B.; Yigitcanlar, T. Pathways to the Making of Prosperous Smart Cities: An Exploratory Study on the Best Practice. *J. Urban Technol.* **2020**, *27*, 3–32. [CrossRef]
46. Rogers, Y.; Marshall, P. *Research in the Wild*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 10. [CrossRef]
47. Chamberlain, A.; Crabtree, A.; Rodden, T.; Jones, M.; Rogers, Y. Research in the Wild: Understanding "in the Wild" Approaches to Design and Development. In Proceedings of the Designing Interactive Systems Conference, DIS'12, Newcastle upon Tyne, UK, 11–15 June 2012; pp. 795–796. [CrossRef]
48. Vavoula, G.; Sharples, M. Future Technology Workshop: A Collaborative Method for the Design of New Learning Technologies and Activities. *Int. J. Comput. Collab. Learn.* **2007**, *2*, 393–419. [CrossRef]
49. Vavoula, G. Future Technology Workshop: A Method for the Design of New Technologies and Activities. 2002. Available online: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.114.1896&rep=rep1&type=pdf (accessed on 1 February 2022).
50. Soro, A.; Brereton, M.; Taylor, J.L.; Hong, A.L.; Roe, P. Cross-Cultural Dialogical Probes. In Proceedings of the First African Conference on Human Computer Interaction, Nairobi, Kenya, 21–25 November 2016; ACM: New York, NY, USA, 2016; pp. 114–125.
51. Gehl, J. *Cities for People*; Island Press: Washington, DC, USA, 2013; ISBN 1597269840.
52. Fainstein, S.S.; Lubinsky, A. The Relationship between Citizen Participation and the Just City: Can More Participation Produce More Equitable Outcomes? In *Learning from Arnstein's Ladder*; Routledge: London, UK, 2020; pp. 129–147. ISBN 0429290098.
53. Monno, V.; Khakee, A. Tokenism or Political Activism? Some Reflections on Participatory Planning. *Int. Plan. Stud.* **2012**, *17*, 85–101. [CrossRef]
54. Gonsalves, K.; Foth, M.; Caldwell, G.; Jenek, W. Radical Placemaking: An Immersive, Experiential and Activist Approach for Marginalised Communities. In *Connections: Exploring Heritage, Architecture, Cities, Art, Media. Vol. 20.1.*; AMPS (Architecture, Media, Politics, Society): Canterbury, UK, 2021; pp. 237–252.
55. Di Bella, A. Smart Urbanism and Digital Activism in Southern Italy. In *Emerging Issues, Challenges, and Opportunities in Urban E-Planning*; IGI Global: Hershey, PA, USA, 2015; pp. 114–140.
56. Anttiroiko, A.-V. City-as-a-Platform: The Rise of Participatory Innovation Platforms in Finnish Cities. *Sustainability* **2016**, *8*, 922. [CrossRef]
57. Desouza, K.; Hunter, M.; Yigitcanlar, T. Under the Hood: A Look at Techno-Centric Smart City Development. *Public Manag.* **2019**, *101*, 30–35.

sustainability

Article

Transformation of Industry Ecosystems in Cities and Regions: A Generic Pathway for Smart and Green Transition

Nicos Komninos

URENIO Research, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; komninos@urenio.org

Abstract: This research paper focuses on pathways towards a digital and green transition. We assess a generic pathway for the transformation of industry ecosystems in cities and regions based on processes of prioritisation, ecosystem identification, and platform-based digital and green transition. We start with problem definition and hypotheses; review related works on transition pathways, such as digital transition, green transition, system innovation, industry ecosystems, and multi-level perspective of transformation; assess the generic pathway with case studies; and conclude with a discussion of findings, outline of conclusions, and policy implications. Overall, the paper investigates pathways, priorities, and methods allowing public authorities and business organisations to master the current industrial transformation of cities and regions introduced by the twin digital and green transitions as an opportunity for radical change of city ecosystems, innovation leapfrogging, and system innovation.

Keywords: industrial transformation; city ecosystem; activity ecosystem; smart ecosystem; pathway; system innovation; digital transition; green transition; smart specialisation

Citation: Komninos, N. Transformation of Industry Ecosystems in Cities and Regions: A Generic Pathway for Smart and Green Transition. *Sustainability* **2022**, *14*, 9694. https://doi.org/10.3390/su14159694

Academic Editors: Tan Yigitcanlar, Bo Xia, Tatiana Tucunduva Philippi Cortese and Jamile Sabatini Marques

Received: 25 June 2022
Accepted: 4 August 2022
Published: 6 August 2022

1. Introduction and Problem Definition

This paper focuses on the ***transformation of industry ecosystems in cities and regions under the influence of digital and green transition***. Industry ecosystems or "activity-based ecosystems" are the most common type of ecosystems in cities, created by companies and organisations that share space, infrastructures, labour market, and other urban externalities. These ecosystems are formed around sectors of economic activity and different vertical markets of manufacturing and services.

There are ecosystems that cities do not choose to develop but grow inherently together with the entire urban system. For instance, cities do not choose to have or not to have housing, transport, energy, and water networks, even though they can choose the type of such ecosystems at the next stage. However, there are ecosystems open to choices, such as the activity ecosystems that flourish in a city. Out of hundreds of different economic activities and industry branches, each city chooses and specialises in a few of them. This choice is evolutionary, based on converging or competing decisions of private and public actors. Nevertheless, the activity specialisation of cities is an outcome of choice.

Our interest in the transformation of activity-based ecosystems by the twin transition, digital and green, is both theoretical and methodological. At the level of theory, we attempt to connect several discrete theory strands dealing with industrial change, innovation, smart systems, and climate-neutral technologies, which are driving this transformation. At the level of methodology, the ambition is to identify pathways to manage the evolving industrial transformation sustained by digital and green transitions. Smart systems and technologies are redefining the industrial landscape. A profound transition of energy systems is underway based on smart energy optimisation and distributed renewable energy production, while climate change adaptation is pushing forward industrial innovation.

This research on pathways of industrial transformation also takes into account the new European policies that appeared after 2020, such as the Green Deal, the new industrial strategy, the policies on digital and green transitions, the research and innovation strategies for smart specialisation, and the good governance of these strategies [1,2]. Overall, this paper investigates pathways, priorities, and methods allowing us to master current industrial transformation as an opportunity for systemic change of cities and regions, innovation leapfrogging, and system innovation introduced by the digital and green transitions.

1.1. Problem Definition

The problem we want to explore concerns the ***pathways of industrial transformation linked to digital and green transitions***. These pathways can (a) connect digital and green technologies enabling a twin transition, (b) produce system innovation leading to a radical change of routines, and (c) transform economic activities and industry ecosystems. Such pathways are of high interest to all countries, regions, and cities. They affect how innovation and transition strategies are implemented and allow for the current transformation of industrial activities and ecosystems to be mastered.

As Figure 1 shows, pathways of change operate in two directions: on the one hand, innovation strategies transform economic activities and their ecosystems through state-led policies, and on the other hand, the twin digital and green transitions transform the same activities through state-led and market-led processes. Therefore, this research can contribute both to research and innovation strategies for smart specialisation (RIS3–RIS4) and the related entrepreneurial discovery processes and to the management of the digital and green transition policies.

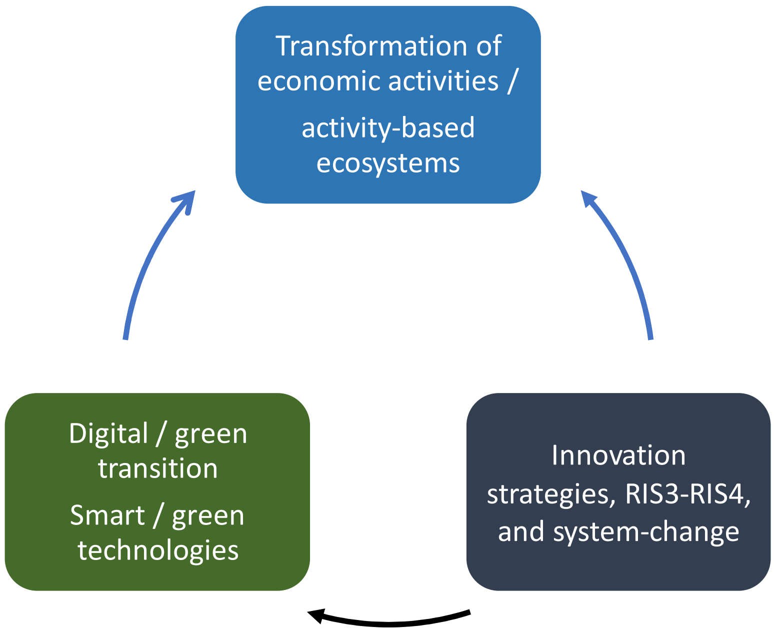


Figure 1. Pathways for innovation, twin transition, and industrial transformation.

The current industrial transformation encompasses all manufacturing, energy, utilities, and services sectors. However, the conditions, technologies, science, and business models of transformation are specific to each sector and change from one sector to the other. The landscape of the current industrial transformation is multifaceted, characterised by different maturity levels [3], different skills [4], and variability of innovation across industries and sectors of economic activity [5].

There is also high variability in pathways of digital and green transitions across economic activities, which are classified by NACE into 88 industry divisions, 272 industry groups, and 615 industry classes [6]. The question is, what economic activities and industries should be placed at the centre of policies for industrial transformation? What ecosystems should a city or region specialise in? How should public funds be invested?

Should all industry divisions, groups, and classes be given equal attention? Are some industry groups more receptive to industrial transformation and effective in high performance, and therefore, should they be placed at the focus of attention?

The sectoral variability of industrial transformation and its pathways is a source of complexity both at management and policy levels. Public authorities and policymakers must elaborate generic pathways for industrial change that can be applied across sectors of economic activity and the numerous vertical markets of industries. We want to find solutions to the "one size does not fit all" problem [7,8].

The search for generic and groundbreaking pathways for industrial transformation is important for many different reasons. It is a global challenge; all counties try to elaborate and promote innovations by smart and green technologies and systems. It is a European challenge, clearly reflected in the new growth policy of the Green Deal, the new industrial strategy of the EU, the research and innovation strategies for smart specialisation, and the strategies for digital transition and ecological transition. Above all, is a challenge for the future of the industry, the future of work, and the well-being of 21st-century societies, cities, and regions.

We want to identify generic trajectories of industrial change with strong potential for a smart transition, high impact on growth and innovation, and minimal environmental footprint or high environmental transition gain. We expect that effectiveness and receptivity to industrial transformation differ with spatial level. Therefore, it is important to combine local, regional, and national perspectives and elaborate pathways of a multilevel government.

1.2. A Generic Pathway towards Smart and Green Transition

The pathway towards a smart and green transition and industrial transformation we want to explore is based on interconnected processes of change that start with smart and green technologies, involve system innovation, and end with the transformation of industrial ecosystems. It is a ***generic transition pathway*** defined by three instances: "prioritisation", "ecosystem perspective", and "platform-based smart and green solutions".

Prioritisation. By focusing on a relatively limited number of important economic activities, even at a high level of granularity, we can capture most economic activities of a city or region. Given that cities and regions tend to specialise in a few industries, prioritising results in dealing with most of the existing industrial activities. Defining pathways of change at the priority level allows for a lowering of complexity in terms of the industrial activities considered for the twin smart and green transitions. It is neither effective nor feasible to go into the details and assess the potential transition pathways of all economic activities in the 272 industry groups or the 615 industry classes of the NACE classification. It is necessary to apply some prioritisation.

The hypothesis here (H1) is that the most important economic activities (by size, specialisation, investment) are expected to include a high share of all economic activities in a territory, and a relatively small number of principal economic activities contain the bulk of all economic activities of a city or region. In other words, there is a high level of polarisation of economic activities in a region.

Ecosystem perspective. In cities and regions, economic activities tend to interconnect, forming activity-based ecosystems. An ecosystem is made by a group of organisations interacting with each other and the environment to achieve common objectives, create value, or other advantages [9,10]. Interconnections with other economic activities occur along supply chains, across vertical markets, over the common infrastructure of cities, as well as in the local labour market and commercial markets. Due to interconnections, activity-based ecosystems are organised and grow.

The hypothesis here (H2) is that we expect to identify ecosystems around the most important economic activities when these are defined by size and specialisation. The size and specialisation increase the probability of interactions and communalities among activities. We expect that activity-based ecosystems occur around economic activities that are large in size and have high specialisation.

Platform-based smart and green transition enables collaborative innovation in products, services, and processes that transform entire ecosystems. On the other hand, ecosystems facilitate the adoption of platforms. Platforms have an impact on many businesses and organisations of an ecosystem and pave the way towards system innovation and the radical transformation of ecosystems. They bring transformations in connections and the organisation overall of the ecosystem, not only its parts, enabling the emergence of routines that change the entire system.

As highlighted in the literature, "industry platforms are technological building blocks (that can be technologies, products, or services) that act as a foundation on top of which an array of firms, organized in a set of interdependent firms (sometimes called an industry "ecosystem"), develop a set of inter-related products, technologies and services" ([11], p. 287). Equally, platforms are described as collaborative business models that sustain ecosystem development: a platform is "a plug-and-play business model that allows multiple participants (producers and consumers) to connect to it, interact with each other and create and exchange value" [12]. Platforms offer the foundation for products and services developed by third parties, a relationship that Gawer and Cusumano [13] name "platform leadership", enabling some companies to exert influence over the direction of innovation in an industry, by engaging other companies to develop complementary products. Platforms are foundations for setting up ecosystems by organisations that share resources, knowledge, or access to markets [14].

The hypothesis here (H3) is that within the economic activities of prioritisation and the ecosystems created around them, we can identify platforms for smart and green transition meaningful for many companies and organisations of the industry or ecosystem. This will enable us to share objectives for collective action, common infrastructure, and pathways for system change for the entire ecosystem to be developed.

The above three instances combine conditions of industry volume and specialisation (prioritisation), systemic organisation (ecosystems), and innovation niches (smart/green solutions). The pathway they define is generic because these instances can be applied in the industry groups of any territory. Hypotheses H1, H2, and H3 measure how representative these instances are within the overall landscape of urban activities and can be assessed with data from case studies.

To do this, the rest of the paper is organised into four sections. In Section 2, we refer to the literature on transition pathways and processes, such as digital transition, green transition, system innovation, industry ecosystems, and multilevel perspective of transformation. In Section 3, we describe a generic pathway that can guide the twin smart and green transitions and assess with case studies how this pathway can be implemented, as well as the outcome of the implementation. In Section 4, we discuss the pathway proposed and the hypotheses described, assessing the feasibility and scenarios of implementation. The last section highlights the conclusions and policy implications.

2. Pathways for Industrial Transformation: Related Works

Previous research on industrial transformation identified multiple drivers, such as digitalisation and smart transition, green transition, system innovation, and ecosystem development [2,15–19]. The arrow of change originates from the twin transitions and moves towards system innovation, ending with the transformation of industry ecosystems. This interconnection is clearly articulated in the new industrial strategy of the European Union, organised around three drivers: a globally competitive and world-leading industry; an industry that paves the way to climate neutrality, the supply of clean and affordable energy and raw materials; and an industry shaping Europe's digital future with investments in artificial intelligence, 5G, and data and metadata analytics [20]. Place-based innovation is also encouraged, allowing regions to develop new solutions with companies and consumers valorising local characteristics, strengths, and specialisations in the framework of smart specialisation strategies [21].

The ***digital (or smart) transition*** is the dominant driver of industrial transformation and refers to the adoption of technologies, such as smart systems, automation and robotisation, sensor networks, Internet of Things, cloud, software, platform and infrastructure as a service (SaaS, PaaS, IaaS, XaaS), analytics, big data, artificial intelligence, and distributed ledger technologies, which transform business environments, operations, and strategies [22–24].

Various terms have been used to describe the current transformation of the industry, such as industrial transition, industry 4.0, smart industry, and the fourth industrial revolution. The digital transition also refers to industries that adopt digital technologies and knowledge-intensive processes. All these terms point towards industrial transformation based on knowledge, information technology, data-based innovation, and a transition from machine-dominant processes to digital.

Industrial transition by digital technologies and smart systems extends to all industry sectors, from agriculture to manufacturing, transport, energy, health, and financial services. Changes to skills and human capital are also associated with the digital transition [25], as well as new business models that connect digitalisation to servitisation and push product companies towards services [26,27]. Industrial transition and industry 4.0 are also characterised by novel processes at the production and enterprise levels, such as smart manufacturing, deployment of embedded actuators and sensors, digital enhancement and reengineering of products, customisation of differentiated products, and coordination of products and services along the supply chain [28]. All these changes require continuous learning and innovation. Overall, the enterprise and industry levels have indeed gathered more attention in terms of research and technology compared with the production level [29].

Assessing the transition to industry 4.0 in the US manufacturing sector, Rojko et al. [30] found that the manufacturing output employment and labour productivity have barely grown. However, the projections for the next decade show brighter developments. They argue that the future will be in cooperation between robots and humans, a partnership that can bring wealth and increase labour productivity, while among the main challenges are the interactions between AI and employees. In this step towards industry 5.0, distributed computers, the Internet of Everything, multiagent systems and technologies, complex adaptive systems, and widespread intelligence are considered the main components of the transition [31]. This new stage in industry development (Industry 5.0) should "focus primarily on human and robot engagement and the integration of human knowledge, creativity, intuition, skills, experience, etc. within robotized production" ([31], p. 303).

The green transition is another major driver of industrial transformation. Guided by the objectives of sustainability and adaptation to climate change, it offers broad opportunities for change due to transversality across industry sectors and territorial scales [22]. Like the digital transition, the green transition has an important systemic impact, as it applies to the entire life cycle of products and engages all segments of a value chain. Systemic for instance is the transition from the "linear economy" of extract, consume, and dispose processes to the "circular economy" that aims to reduce, reuse, and recycle.

Geels [15] investigated the fundamental changes in energy, transport, housing, and agrofood systems related to sustainability. He identified different types of innovations in energy and transport systems, including radical technical innovation (battery electric vehicles, decarbonisation), grassroots and social innovation (car sharing, bike clubs), and business model innovation (mobility services and infrastructural innovation (intermodal transport, compact cities)). Zhai and An [32] analysed the factors influencing the green transformation in China's manufacturing industry with a survey of 500 Chinese enterprises and identified human capital, financing strength, technology innovation, and government policy as having a significant positive impact on the green transformation performance. Governmental behaviour had the greatest impact coefficient, followed by human capital, technology innovation, and financing ability. On the contrary, environmental regulation decreased the positive impact and acted as a reversal mechanism affecting financing capacity, technology innovation, and governmental behaviour.

The European Green Deal is also expected to make an important contribution to the green transition. It is the new growth strategy for the European Union and an integral part of implementing the United Nations 2030 Agenda and the Sustainable Development Goals. The EU Green Deal is holistic and covers all areas of activity, climate, energy, transport, industry, construction, and nature. In response to these challenges, it outlines a development strategy to transform the EU into a just and prosperous society, with a modern, resource-efficient, and competitive economy, without greenhouse gas emissions in 2050 and economic growth decoupled from resource use [1,33–35].

It is important to underline that the green transition together with efforts for renewable energy and CO_2 reduction promotes processes of reuse, zero waste, and modular production that allows repair and replenishment rather than total rejection of products. Through reuse, the green transition converges with the digital transition in the continuous reuse of knowledge products. Knowledge, as shown by the new growth theory, is not only not consumed during use but is improved by repetition and reuse [36,37]. Both the digital transition and the green transition are based on a wide range of technologies, systems, and solutions.

System innovation or transformative innovation is a direct outcome of radical changes introduced by the digital and green transitions. Already, the term "transition" brings in the idea of movement or change from one state of a system to another. This type of innovation goes beyond products and technologies and involves changes in the broader sociotechnical system. System innovation is characterised by large-scale transformations having wide societal value, such as energy, housing, mobility, and food; transformations through coevolution between different elements and actors; and transformations that occur at multiple levels, such as the niche level, the regime level, and the landscape or wider political and economic level [38]. This is a new framing of innovation that emphasises system-level changes in the structure or architecture of the system of reference [39–43].

This type of innovation encompasses both production and consumption activities and the complex relationships of actors ranging from firms and knowledge producers to households and consumers. Government has a more important role through policies enabling system-level innovations. As Pontilakis et al. [2] point out, system-level innovations do not have a single designer and are codeveloped through countless contributions within industry ecosystems. Therefore, distributed agency and being loosely connected by fleetingly aligned interests are key features, as well as the identification of interconnections between disparate parts of a system and potential domains for policy intervention, in particular, interventions for radical change through smart specialisation strategies.

In less developed regions, system innovation may have an important leapfrogging effect. For instance, environmental leapfrogging can enable developing countries and regions to skip some of the "dirty" stages of development followed in the industrialised world and contribute to environmental goals and climate mitigation solutions [44–46]. The same holds true for industry 4.0, where leapfrogging innovation can offer momentum in the dynamics of industrial growth with the early adoption of advanced digital systems [47–49].

The turn towards industry ecosystems is another important new dimension of the current industrial transformation driven by the twin digital and green transitions. It highlights a change of focus from individual companies to groups of organisations connected at multiple spatial scales [50,51]. An industry ecosystem is an organic network of collaboration among two or more business entities that create and share assets and value. It must be distinguished from an innovation ecosystem, which refers to organisations (R&D, producers, financiers, market makers) that collaborate in new product development and innovation. Industry ecosystems appear as global manufacturing networks [52], cross-industry ecosystems [17], platform ecosystems [53–55], and local entrepreneurial ecosystems [56].

In a review of the ecosystem concept in the field of management, Tsujimoto et al. [10] provide an overview of 90 studies that use the concept and identify four types: *industrial* *ecosystems* based on the industrial ecology perspective, material and energy flows, and interaction with the environment; *business ecosystems* based on the theory of organisational boundaries, comprising digital ecosystems, cross-industry ecosystems, supplier ecosystems, and business group ecosystems; *platform ecosystems* organised in two-sided markets; and *multiactor network ecosystems* based on social network theory.

The shift to ecosystems is clearly articulated in the updated EU industrial strategy that identifies 14 industry ecosystems as being important for the EU, including aerospace and defence; agrifood; construction; cultural and creative industries; digital; electronics; energy-intensive industries; energy renewables; health; mobility, transport, automotive; proximity, social economy, civil security; retail; textiles; and tourism [57].

The multi-level perspective (MLP) offers a theoretical framework that allows integrating the above-mentioned elements of industrial transformation, twin transitions, system innovation, and industry ecosystems. The MLP was developed by Rip and Kemp [58] and was further elaborated and refined by Geels [59] and Geels and Schot [60]. It is an attempt to bring together different strands of innovation theory, such as evolutionary economics, the sociology of innovation, neoinstitutional theory, and science and technology studies, and combines overlapping but disconnected themes of technological change and innovation [61].

The MLP focuses on radical innovations or system-change innovations. These are enacted by the tandem action of multiple social groups, enterprises, consumers, social movements, policymakers, researchers, media, and investors. In this sense, the MLP comes closer to quadruple helix innovation perspectives. Geels [15] points out that the MLP as a process theory "has both a 'global model' component (consisting of three analytical levels and several temporal phases), which describes the overall course of socio-technical transitions, and a 'local model' component, which addresses-specific activities and causal mechanisms in multi-level interactions". Transformations are nested at three levels within the system, the landscape (macrolevel), the regime (mesolevel), and the niche (microlevel). The theory gives more emphasis on the role of agency and transition pathways to new states of a system [62].

A system-level transition starts when the prevailing sociotechnical regime shows significant problems, key innovations appear that drive new designs, and early adoptions of the transition technologies take place. Geels and Schot [60,63] have identified five transition pathways: (a) the transformation of sociotechnical regimes without recourse to one dominant technology, (b) technological substitution when a radical technology replaces an existing technology, (c) dealignment and realignment of existing regimes when competing for new technologies solving existing problems, (d) opening up a new sociotechnical system that offers new social functions, and (e) reconfiguration and system change when many technologies and organisations change.

More recent works have connected MLP with smart specialisation strategies and technological changes in local industrial systems, considering MLP as a place-based driver for the technological transition of regional economies [64,65]. De Propris and Bailey [16] suggest that the transformation of a local system rests on three types of capabilities: innovation capabilities, docking capabilities to attract delocated niches, and translational capabilities to absorb radically new technologies. They identify four transformative pathways—(a) endogenous, (b) hypertransformative, (c) importation, and (d) regional obsolescence—and argue in favour of a transformative place-based policy enabling the joining up of technologies, sectors, and places, through a transformative entrepreneurial discovery process.

3. Towards a Generic Pathway of Transition: Evidence from the Case Studies

3.1. Generic Pathway Instances and Hypotheses from a Multilevel Perspective

In the MLP approach, two branches of research on transitions can be identified, referring to systems in transition and management of the transition. This distinction indicates an analytical versus an interventionist approach that focuses on how to actively steer technological change and how purposive, science, and technology-led transitions can be organised [61,66,67].

The generic pathway we described in Section 1.2, its three instances, and related hypotheses (prioritisation, ecosystem perspective, platform-based smart and green transition) stem from the above understanding of transition as system changes in cyber-physical systems of innovation. Due to digital transition, the continuous widening of digital networking, rich and real-time data availability, e-infrastructures, and e-services, all innovation systems are currently becoming cyber-physical. Their physical and institutional dimensions are interwoven with a strong digital dimension (Figure 2).

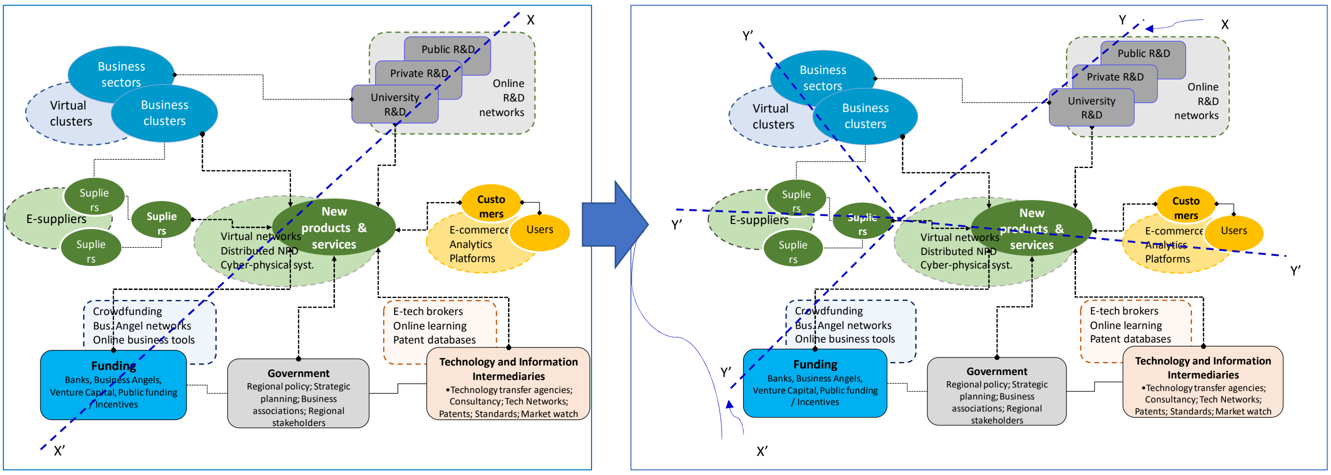


Figure 2. System innovation in cyber-physical innovation systems.

The "prioritisation" is mainly justified by the absence of theoretical prediction on how industries are affected by the twin digital and green transitions. It is highly probable to find innovative solutions in less expected economic activities. Therefore, all 272 NACE industry groups should be reviewed as potential fields of promising transition, which demands an enormous effort from policy-making authorities. Prioritisation is also a cornerstone of smart specialisation strategies and the entrepreneurial discovery process (EDP). As has been noted, "Smart Specialisation should address the difficult problem of prioritisation and resource allocation based on the involvement of all stakeholders in a process of entrepreneurial discovery, which should secure a regionally and business-driven, inclusive and open prioritisation process" [68]. Thus, prioritisation allows for transformations at the landscape macrolevel to be managed, giving priority to certain industry groups. It is meaningful if hypothesis H1 is valid, and an important share of economic activities is included in the selected priority activities.

The "ecosystem perspective" is also strongly related to the twin transition, as digital and green transitions initiate system innovations that change the entire networking architecture of ecosystems, not just products and services [69]. However, there is something more. The current dominant pathway for product and service innovation is based on close associations between "research breakthroughs", "venture capital funding", and "startup creation" along connectivity illustrated as X-X′ axis in the cyber-physical system (Figure 2). In system innovation, this pathway and the networking architecture change, and the focus moves from startups to supply chains along with wider networking, illustrated as Y-Y′. The twin transition moves the entire system from state A to a new state B. For instance, the green transition combined with digital collaboration in the energy ecosystem introduces renewable energy, energy optimisation, and nature-based solutions, and changes the entire energy ecosystem, not only the innovative products of startups. The ecosystem perspective allows transformation to be organised at the regime mesolevel, connecting actors of the industry, science, technology, consumers, and policymakers into a new regime. It is meaningful if H2 is valid, and within the selected priority economic activities, we find a strong presence of ecosystems.

The "platform-based smart and green transition" allows transition to be organised at the niche microlevel, enabling an important number of actors of an ecosystem to adopt innovative solutions in tandem. As mentioned, ecosystems make easier the adoption of platforms. Over platforms, niche actors, entrepreneurs, startups, and spinoffs can experiment with radical innovations that deviate from existing regimes and propel the entire system towards a new state.

3.2. Evidence from the Case Studies

We assessed this generic pathway of industry transformation in research we conducted for the European Commission, DG Regional and Urban Policy, titled "Ecosystems and functioning EDP for S3 2021–2027" [70–72]. We investigated pathways of industry change in Greece and Cyprus relevant for research and innovation strategies for smart specialisation. The research was placed in the framework of good governance of national and regional smart specialisation strategies 2021–2027, which is assessed by seven fulfilment criteria, among which is the "functioning of stakeholder cooperation in the entrepreneurial discovery process".

The main rationale of EDP within RIS3 is that European regions should explore and exploit key capabilities for global niche markets and create long-term competitive advantages [21,73–75]. EDP is expected to reveal innovative, but place-specific and evidence-based, opportunities that take advantage of available resources and competencies. During the EDP, different entrepreneurial actors are brought together in a government-led participatory process generating a collective debate, integrating the divided and dispersed knowledge belonging to different actors, and setting common priorities for intervention.

Thus, the objective of the EDP is to identify pathways for industrial diversification and transformation towards higher added value activities [68]. Diversification may be *intraindustry*, when research and innovation change and improve the products and processes of an industry, or *interindustry*, when innovation leads to a branching of an industry towards other sectors. Interindustry diversification may be "*related*" or "*unrelated*" to existing skills and know-how. Empirical evidence suggests that knowledge spillovers within a region occur primarily among "related" economic activities and only to a limited extent among "unrelated" ones [76]. It is the "related variety" in a region that feeds branching out to new activities from technologically related activities, not regional diversity or regional specialisation *per se* ([77], p. 67). Unfortunately, we do not have any theoretical guidance about the diversification of industries in the other trajectories, in the case of either an intraindustry unrelated change or an interindustry unrelated change.

This theory gap is accompanied by a methodology gap regarding the granularity of the EDP. The granularity allows the level of detail to be defined when modelling industries or decision-making processes. The greater the granulation, the deeper the level of detail and the better the understanding of future trends. However, we do not proffer any methodological guidance about the best industry granularity level to perform the EDP. For instance, is it better to perform the EDP at the level of industry sections, industry divisions, industry groups, or industry classes?

Given these gaps, we addressed the functioning EDP as a transformation pathway defined by "prioritisation", "ecosystem identification", and "platform-based innovation", first, by identifying the most important economic activities per region; second, by identifying ecosystems, intra- or inter-regional, that have the potential for future growth and inclusive growth for most of their members; and third, by evaluating the potential for innovation, especially platform-based innovation and smart/green transition. Consequently, this research was conducted in three consecutive stages.

Stage 1: Identification and prioritisation of economic activities. As mentioned, NACE Rev. 2 classifies economic activities in industry sections, divisions, groups, and classes. Regional data are available for sections, divisions, and groups, and in some cases, for industry classes. Usually, the industry group level is at the level of higher granularity and detail when it comes to regional data. If the EDP is manageable at this level of detail, then the industry group level is preferable to any other level of granularity.

Data on the regional distribution of industry groups in Greece is provided by ELSTAT. We used the dataset of 2017. In this dataset, three variables are given per region and industry group: (1) number of legal entities (companies), (2) turnover, and (3) number of employees. Based on this dataset, we calculated two more indicators: (4) the location quotient based on the number of companies and (5) the location quotient based on the number of employees. The location quotient allows for the strength and size of a particular industry in a region to be evaluated. It quantifies how concentrated an industry is within an area compared with the country as a whole. It is the most preferred index of specialisation, calculated as a proportion of an industry in a region compared with the proportion of the same industry in the country. Having those five variables, we created our basic data matrix, which comprised 7 columns and 3536 lines (272 industry groups $\times$ 13 regions).

For each one of the above five variables, we ordered the industry groups per region and selected the top 10 by size and specialisation. We produced four ordered lists, by the number of companies, the number of employees, the location quotient on companies, and the location quotient on employment (top 40 industry groups). Then, we cleaned these ordered lists by removing industry groups with limited entrepreneurial activity, such as public companies, utilities provided by public authorities, public services for administration, defence, libraries and museums, and service sectors in which self-employment dominates, legal, accounting, veterinary, and so on.

Per region, the ordering and cleaning of industry groups by size (number of companies and employment) and specialisation (location quotient on the number of companies and employment) produced a list of the ***top 40 groups***, in total, 520 industry groups in the 13 regions of Greece. However, this was not a combined ordering. To arrive at a combined ordering of industry groups per region, we selected one after the other, industry groups at the top 10 positions in all four lists, industry groups at the top 10 positions in three out of four lists, industry groups at the top 10 positions in one list related to size and one list related to specialisation, and industry groups in two lists of specialisation. Table 1 shows the logic for selecting the top 10 industry groups per region. We start with the selection of groups that figure in all lists of size and specialisation and move down to industry groups of high specialisation.

Table 1. Selection of top 10 industry groups per region.

Top10 per Number of Companies		Top10 per Employment		Top10 per Specialisation on Companies		Top10 per Specialisation on Employment	
NACE	Index	NACE	Index	NACE	Index	NACE	Index
55.1	1077	55.1	20,284	10.4	8.12	10.4	6.39
10.7	591	10.7	3241	30.3	4.55	55.1	2.97
72.1	499	79.1	2570	23.4	3.55	50.1	2.80
10.4	466	50.1	1707	32.2	3.18	23.4	2.30
79.1	378	72.1	1323	55.1	2.51	79.1	2.04
62.0	351	10.4	1237	79.1	2.14	72.1	1.69
90.0	269	10.1	791	72.1	1.98	32.2	1.65
31.0	235	31.0	699	25.2	1.88	13.9	1.51
16.2	208	62.0	663	10.5	1.87	10.1	1.14
10.5	95	61.2	624	16.2	1.79	28.3	1.13

NACE	Top10 per Number of Companies	Top10 per Employment	Top10 per LQ on Companies	Top10 per LQ on Employment
55.1	1077	20,284	2.51	2.97
10.4	466	1237	8.12	6.39
72.1	499	1323	1.98	1.69
79.1	378	2570	2.14	2.04
16.2	208		1.79	
50.1		1707		2.80
10.1		791		1.14
10.5	95		1.87	
23.4			3.55	2.30
32.2			3.18	1.65

We consider these industry groups as the most important industry groups per region because they exhibit both large size and high specialisation. Looking at all 13 regions together, we find that the top 10 industry groups belong to 51 categories only, of which 26 categories appear in more than one region and 25 in one region only (Table 2). The 26 interregional industrial groups hold 105 of the 130 (81%) positions in the top 10 industries in all regions of Greece. From a prioritisation perspective, this finding shows that in 51 industrial groups, we can explore the most important economic activities in Greece, while 26 industrial groups capture 81% of the most important economic activities in the country.

Table 2. Most important (top 10) industry groups in regions of Greece.

NACE	Name	Number of Regions	NACE	Name	Number of Regions
55.1	Hotels and similar accommodation	8	63.9	Other information service activities	1
11.0	Manufacture of beverages	8	61.3	Satellite telecommunications activities	1
10.5	Manufacture of dairy products	7	61.1	Wired telecommunications activities	1
03.1	Fishing	7	50.2	Sea and coastal freight water transport	1
16.2	Manufacture of products of wood, cork, straw, and plaiting materials	6	32.2	Manufacture of musical instruments	1
31.0	Manufacture of furniture	5	32.1	Manufacture of jewellery, bijouterie, and related articles	1
03.2	Aquaculture	5	30.3	Manufacture of air and spacecraft and related machinery	1
25.1	Manufacture of structural metal products	4	29.1	Manufacture of motor vehicles	1
23.4	Manufacture of other porcelain and ceramic products	4	28.9	Manufacture of other special-purpose machinery	1
10.9	Manufacture of prepared animal feeds	4	26.7	Manufacture of optical instruments and photographic equipment	1
10.7	Manufacture of bakery and farinaceous products	4	26.2	Manufacture of computers and peripheral equipment	1
10.6	Manufacture of grain mill products, starches, and starch products	4	26.1	Manufacture of electronic components and boards	1
10.3	Processing and preserving of fruit and vegetables	4	24.3	Manufacture of other products of first processing of steel	1
90.0	Creative, arts, and entertainment activities	3	24.2	Manufacture of tubes, pipes, hollow profiles, and related fittings of steel	1
79.1	Travel agency and tour operator activities	3	23.6	Manufacture of articles of concrete, cement, and plaster	1
72.1	Research and experimental development on natural sciences and engineering	3	23.3	Manufacture of clay building materials	1
50.1	Sea and coastal passenger water transport	3	22.2	Manufacture of plastic products	1
23.7	Cutting, shaping, and finishing of stone	3	21.1	Manufacture of basic pharmaceutical products	1
16.1	Sawmilling and planning of wood	3	20.5	Manufacture of other chemical products	1
10.4	Manufacture of vegetable and animal oils and fats	3	18.2	Reproduction of recorded media	1
10.2	Processing and preserving of fish, crustaceans, and molluscs	3	15.1	Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, and harness; etc.	1
10.1	Processing and preserving of meat and production of meat products	3	14.2	Manufacture of articles of fur	1
62.0	Computer programming, consultancy, and related activities	2	14.1	Manufacture of wearing apparel, except fur apparel	1

Table 2. *Cont.*

NACE	Name	Number of Regions	NACE	Name	Number of Regions
28.3	Manufacture of agricultural and forestry machinery	2	13.3	Finishing of textiles	1
22.1	Manufacture of rubber products	2	10.1	Processing and preserving of meat and production of meat products	1
10.8	Manufacture of other food products	2			

White for the primary sector, green for manufacturing, and brown for services.

Stage 2: Identification of business ecosystems. It is an important finding that 51 industry groups, which gather activities at a high granularity level, capture the most important economic activities of a country. Now, at stage 2 of research, we moved further and searched for ecosystems within those 51 industry groups.

This survey was carried out in 2020 and covered all 13 NUTS 2 regions of Greece. It was based on field research and interviews with companies and experts from relevant stakeholders and agencies to trace value chains, common strategies, common infrastructures, or operating platforms among the companies in the top 10 industry groups of each region. We prepared 13 questionnaires allowing us to identify the three most important business ecosystems per region. An example can be found at https://www.surveymonkey.com/r/7FKJVHF (accessed on 26 June 2022).

The survey showed that among the 51 identified industry groups, 25 have ecosystem features. They share a common infrastructure, natural or energy resources, or technology; they work with common platforms or are part of the same value chain. Moreover, these industry groups have typical characteristics of business complexes, such as geographical boundaries in one area, productive specialisation, and location quotients higher than 2 in all cases and higher than 10 in some cases. Those 25 industry groups/ecosystems are listed in Table 3. Most ecosystems are interregional, indicating the need for multilevel government across cities and regions.

Table 3. Key features of identified ecosystems/industry groups.

REGION	Industry Group/Ecosystem	Size of Ecosystem	Mature/ Emerging	R&D and Innovation Demand	Innovation Platform	Regional/Interregional
East Macedonia and Thrace	22.2 Manufacture of plastics	Small	Mature	Medium	New product and materials	Regional
	23.7 Cutting, shaping of stone	Large	Mature	Medium	Brand and by-products	Interregional
	26.2 Manufacture of computers	Small	Emerging	High	No	Regional
Central Macedonia	10.3 Processing fruit and vegetables	Large	Mature	High	Brand and packaging	Interregional
	14.1 Manufacture of wearing apparel	Large	Mature	Medium	Brand and design	Regional
	25.1 Manufacture of structural metal products	Large	Mature	Medium	Materials	Regional

Table 3. *Cont.*

REGION	Industry Group/Ecosystem	Size of Ecosystem	Mature/ Emerging	R&D and Innovation Demand	Innovation Platform	Regional/Interregional
West Macedonia	16.2 Manufacture of products of wood	Large	Mature	Low	Brand and eco-quality	Interregional
	14.2 Manufacture of fur	Large	Mature	Low	Export	Regional
Epirus	10.1 Processing of meat	Medium	Mature	Medium	Brand and packaging	Interregional
	10.5 Manufacture of dairy products	Large	Mature	High	Brand and packaging	Interregional
Thessaly	22.1 Manufacture of rubber products	Small	Emerging	Low	No	Regional
	31.0 Manufacture of furniture	Large	Mature	Low	Commercial infra	Interregional
Sterea Ellada	24.2 Manufacture of tubes of steel	Small	Mature	Low	New product	Regional
Ionian Islands	79.1 Travel and tour operator activities	Large	Mature	High	New product	Interregional
Attica	90.0 Creative, art activities	Large	Mature	High	Digital infrastructure	Interregional
	62.0 Computer programming	Large	Emerging	High	Market and infrastructure	Regional
	21.1 Manufacture of pharmaceutical products	Small	Emerging	High	New product	Regional
Western Greece	03.2 Aquaculture	Medium	Mature	Medium	Brand and new product	Interregional
	10.9 Manufacture of prepared animal feeds	Medium	Mature	Medium	Production and supply chain	Interregional
Peloponnese	11.0 Manufacture of beverages	Large	Mature	High	Production and by-products	Interregional
North Aegean	10.4 Manufacture of vegetable oils and fats	Large	Mature	High	Brand and quality	Interregional
	03.1 Fishing	Large	Mature	Low	Brand and Infrastructure	Interregional

Table 3. *Cont.*

REGION	Industry Group/Ecosystem	Size of Ecosystem	Mature/ Emerging	R&D and Innovation Demand	Innovation Platform	Regional/Interregional
South Aegean	50.1 Sea passenger water transport	Large	Mature	Low	Infrastructure	Interregional
Crete	55.1 Hotels and similar accommodation	Large	Mature	High	Market access	Interregional
	72.1 Research in natural sciences and engineering	Large	Emerging	Medium	Infrastructure	Interregional

Stage 3: Opportunities for platform-based digital and green transition. At this third stage of research, we further studied the 25 ecosystems identified, sketching their profile, assessing bottlenecks for innovation, needs and demand for innovation, and potential platforms that can lead to their smart and green transformation. Areas of ecosystem diversification were explored to better understand emerging trends and future growth areas.

The survey was based on secondary data from various sources, such as sectoral studies published by the Foundation for Economic and Industrial Research (IOBE) or other industry organisations; data from business directories on financial performance per industry and other secondary data sources, such as company websites, news, and reports from industry associations; and data from research proposals submitted in response to two national calls for research and innovation support. A report was prepared for each of the 25 ecosystems providing information on the ecosystem profile, relationship to regional research and innovation policy priorities, business and growth challenges, research and innovation demand, common challenges, and potential areas for platform-based ecosystems.

Based on this information, we produced a typology of the 25 ecosystems combining size, business challenges, and innovation demand, which reveals four different types of ecosystems, clustered around challenges of product design and development, production and supply chain optimisation, branding and promotion, and market innovation and export market access (Figure 3).

New product design and development is the dominant innovation challenge in ecosystems, such as 21.1–manufacture of basic pharmaceutical products (new medicines and molecules, pharmaceutical discovery, relocation and drug retargeting), 22.2–manufacture of plastic products (new degradable plastics, transition to a circular model), 55.1–hotels and similar accommodation (services to specific population targets, digital applications to provide advanced services or optimise existing services), 62.0–computer programming and consultancy (smart applications and new e-services), 79.1–travel agency and tour operator activities (replacement of services previously offered, design of new services). This challenge is pertinent for large and small ecosystems; emerging ecosystems, such as pharmaceuticals; or mature ecosystems, such as hotels and accommodation.

Production modernisation, supply chain optimisation, and environmental sustainability is the dominant innovation challenge in ecosystems, such as 03.2–aquaculture (improving the productivity, diagnosis and control of diseases, expansion of activities), 10.1–processing and preserving of meat and production of meat products (verticalisation, standardisation and processing, storage and distribution), 10.9–manufacture of prepared animal feeds (increased specialisation, supply of raw material, lowering production costs), 11.0–manufacture of beverages (protocols for the clonal selection of grapevine, vertical coordination, high labour costs), and 23.7–cutting, shaping, and finishing of stone (automation, exploitation of mining and marble by-products, environmental remediation, quarry rehabilitation). These innovation challenges are pertinent for large and medium-size ecosystems, characterised by midlevel demand for research and innovation and needs for technology transfer rather than radical process innovations.

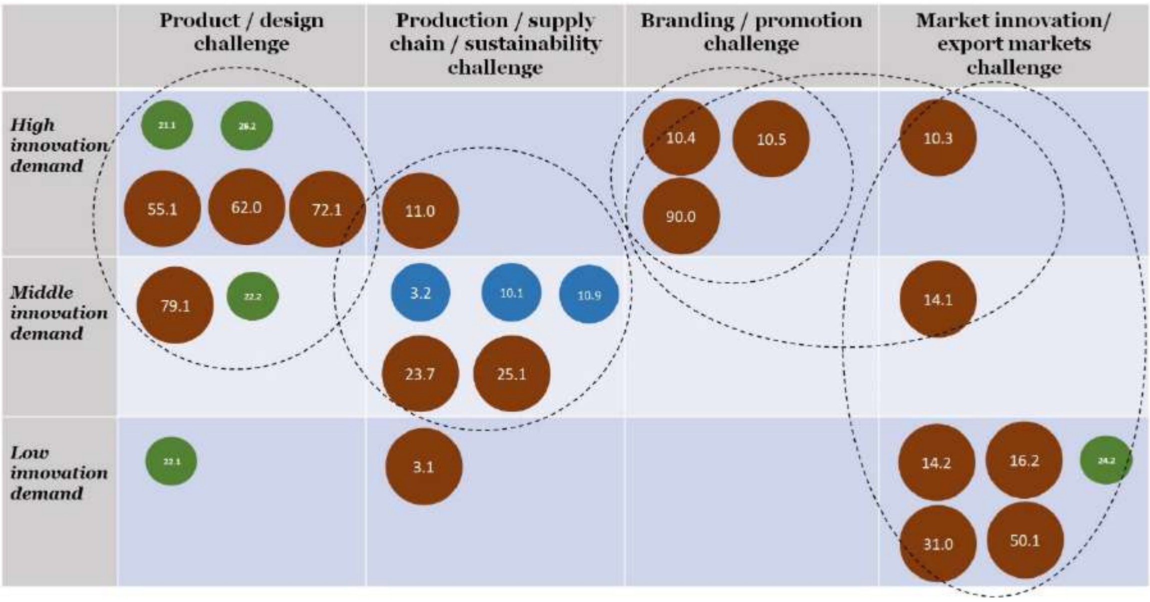


Figure 3. Areas for platform-based innovation in the 25 identified business ecosystems (brown, ecosystems with more than 200 companies; blue, ecosystems having between 50 and 200 companies; green, ecosystems with less than 50 companies).

Branding and promotion are the dominant innovation challenges in ecosystems, such as 10.4–manufacture of vegetable and animal oils and fats (high quality of products but low branding, standardisation of quality, trade-in bulk form), 10.5–manufacture of dairy products (local brands, better packaging, international sales networks), 90.0–creative, arts, and entertainment activities (access to media, innovative platforms for promotion, dissemination of intangible cultural heritage).

Market innovation and access to global markets is the dominant innovation challenge in ecosystems, such as 4.2–manufacture of articles of fur (sharp drop in demand from abroad, lost market shares due to traditional promotion models). In the internal market, the collapse of demand due to the construction sector crisis exerts pressure in industries, such as 16.2–manufacture of products of wood, cork, straw, and plaiting materials; 24.2–manufacture of tubes, pipes, hollow profiles, and related fittings; 31.0–manufacture of furniture, making urgent the turn towards new markets. The 50.1–sea and coastal passenger water transport was also affected by the crisis. All these ecosystems are mature, characterised by low-level innovation capabilities and demand. This is an additional barrier to industrial transformation.

The profiles of industry groups/ecosystems also reveal the potential for platform-based development to address common challenges of companies belonging to an ecosystem. We identified product, production, trade, technology, and environmental challenges, and consequently, platforms were identified in 23 cases to lead the twin digital and green transitions. Platforms may be physical, institutional, infrastructural, or digital. They can be *market-driven*, providing access to markets, branding, and promotion; *product-driven* for new product design and development, smart products, product quality, and certification; *technology-driven* to facilitate research, processing technologies, and supply chain integration/optimisation; *infrastructure-driven* to provide physical, institutional, and digital infrastructure; and *materials-driven* to better manage new materials, raw materials, waste, and recycling. Such platforms strengthen the ecosystems identified, acting as anchors for orchestrating complementors.

Technologies to be used in platform development are listed in Table 4. These are smart and green technologies to be applied at the company and ecosystem levels, enabling the orchestrated innovation and growth of the respective ecosystem.

Table 4. Technologies for digital and green transition.

	Smart Technologies	Green Technologies	Smart–Green Technologies
Company level	• ERP, CRM • e-Commerce • Digital marketing • Automation • IoT, smart meters • AI • Data and analytics	• Circular design • Waste treatment • Recycling • Renewable energy (RE) • Energy storage • Energy saving • Building retrofitting	• Energy optimisation • Energy saving • Materials optimisation • Telework • Digital twins
Ecosystem level	• Branding • Two-sided platforms • Marketplace e-commerce • Crowdsourcing • Supply chain optimisation • Cloud, smart infrastructure • Data and analytics	• The above plus • Large-scale RE • Ecosystem-based RE storage • Energy communities • Footprint benchmarking • Nature-based solutions	• Energy sharing platforms • Smart grid • Smart grid energy storage • Data dashboards • Pollution alert • Digital twins • Blockchain self-organisation

A good working example of platform-based innovation is Mediterra S.A, the research and innovation centre of the mastiha producers of Chios Island. It was founded by the Chios Mastiha Growers Association for product development and marketing of mastiha and the promotion and sales of mastiha products worldwide. To date, the company has developed a retail outlet network under the brand "mastihashop" that comprises stores in Greece and abroad, has established a food production facility in Chios where over 100 different products are produced, and has developed a wide distribution network for brands, such as natural mastiha, mastiha chewing gum, cosmetic products, parapharmaceutical products (selling line: mastiha therapy), and Greek food products (selling line: cultura mediterra). The centre performs R&D on the antibacterial activity of mastiha, nonoxidative action, mastiha in oral hygiene, dermatological and healing properties of mastiha, and new product development using mastiha as a natural supplement to functional foods. Own facilities cover an area of approximately 10,000 m^2 and house the total range of activities, including two production units for mastiha processing and packaging, testing new products, and distillation of mastiha oil.

Another example is a smart–green platform for the industry ecosystem, 10.3–processing and preserving of fruit and vegetables, which brings together companies from Central Macedonia (177), Western Greece (64), Thessaly (74), and Peloponnese (78) with 9601 employees and EUR 1.382 billion turnover (2017). This platform promotes green production and nonplastic packaging, which is a common challenge among companies in this industry. The aim is to create a high-quality brand that provides also quality certification, branding products for green production and alternatives to plastic packaging. Demand for sustainable production and packaging is likely to increase during the next years, and their early adoption can provide a competitive advantage for fruit producers. Using digital tools, the platform offers to all participating companies' full information and traceability of products throughout the supply chain. At the same time, the platform can work as a competence centre promoting learning and the adoption of green production technologies and related smart systems in the processing of agricultural products. This may further enhance the competitiveness of this transregional ecosystem of processing and preserving fruits and vegetables.

4. Discussion

The literature on the current industrial transformation in cities and regions reveals the central role of smart and green technologies in enabling system innovation or transformative innovation through the processes of digitalisation, optimisation, dematerialisation, CO_2 reduction, and circularity. The multilevel perspective offers a good theoretical framework that allows industrial transformation, twin transitions, system innovation, and industry ecosystems to be connected and integrated. The interest in industry ecosystems and platforms, enabling the formation of ecosystems, is a direct outcome of system-level changes that transform the organisation of industries, not only their products and services.

The contribution of the present paper to this debate is through the assessment of a generic pathway for managing the transformation of activity ecosystems in cities and regions, which stands on instances of "prioritisation", "ecosystem identification", and "platform-based smart and green transition". The case studies we summarily presented provide good feedback on the feasibility of this generic pathway and how its three instances work together and complement each other.

We have seen that prioritisation with respect to size and specialisation allows the complexity of industrial transformation to be lowered. At a level of high industry granularity, instead of considering the transformation of 272 industry groups, we can focus on 51 groups only. In Greece, these top 10 industry groups per region capture an important share of industrial activity, including 34.04% of companies, 38.57% of employment, and 42.20% of turnover. In the Cyprus case study, also working with top 10 industry groups by the number of companies, size of employment, production value, fixed capital investments, and emerging industries, the same prioritisation method allowed us identify 16 industry groups that account for the lion's share of the overall industrial activity, including 43.33% of companies, 57.37% of employment, 64.34% of production value, and 72.73% of fixed capital investment.

Prioritisation and a focus on a smaller collection of industry groups pave the way for surveys on ecosystem identification. Within the 51 top 10 industry groups across the 13 regions, we identified 25 ecosystems (see Tables 2 and 3). Most ecosystems are large (17), having more than 200 companies; fewer are small (5); and even fewer are midsized ecosystems (3). Additionally, the majority are established mature ecosystems, justifying a deviation from the startup innovation model towards a model engaging existing supply chains and wider networking, as illustrated in Figure 2.

The third stage of the case studies focused on identifying technologies and platforms that offer opportunities for digital and green transition. Platforms, on the one hand, upgrade products, services, and processes at the ecosystem level and, on the other, affect a large number of businesses and organisations that are active in the ecosystem. Working with industry-wide platforms involves a two-part structure: on the one side is the platform with its infrastructure, hardware, software, and data, and on the other side are the organisational or technological solutions hosted on the platform. A typology proposed by Srnicek [78] classified platforms according to their purpose: *advertising platforms* that offer advertisement space, *cloud platforms* that offer hardware and software as a service, *industrial platforms* that offer infrastructures for the transformation of manufacturing, *product platforms* that generate revenue by offering goods as a service, and *lean business model platforms*. In platform-based ecosystems, the orchestration of producers and consumers is achieved by the platform, its services and infrastructures, and the business model for viability. Platforms offer services or infrastructure and have income from these services that ensure their sustainability.

All three hypotheses related to the instances of the generic pathway for industry transformation have been verified by the case studies: H1, that most important economic activities (by size, specialisation, investment) have a high share of all economic activities and a relatively small number of principal economic activities account for the mass of all economic activities of a city or region; H2, that we can identify ecosystems around the most important economic activities when these are defined by size and specialisation; and H3, that within the economic activities of prioritisation and the ecosystems created around them, we can identity platforms for smart and green transition relevant for many companies and organisations of the industry or ecosystem.

Working along with these three instances that define a generic pathway for industry transformation, the critical path is related to the third instance of platform-based transition. Platforms providing services for market making (access, branding, promotion), product development (innovation, quality, certification, standardisation), and technology development (materials, processing, supply chain optimisation, circularity) are mostly needed to collectively address the innovation and transformation challenges faced by activity ecosystems. They give birth to or strengthen ecosystems created around common challenges. Platforms and ecosystems also guarantee the public character of the innovation policy as they serve the common needs of industry groups rather than the interests of individual companies in the group. The collective character of innovation and transformation trajectories is introduced by user and stakeholder engagement in decisions about platform selection, deployment, and operation procedures. This is a standard procedure within smart environments [79,80].

5. Conclusions

In this paper, we described and assessed a generic pathway for managing the transformation of activity ecosystems in cities and regions defined by the processes of "prioritisation", "ecosystem identification", and "platform-based digital and green transition". These three processes drive a system change of ecosystems, as outlined by the multi-level perspective in the socioeconomic landscape (wider trends of globalisation, population, financial conditions, lifestyles), sociotechnical regime (conventional routines and rules), and niches (new technologies and practices) [81].

The three instances of this generic pathway work in tandem. "Prioritisation" lowers complexity and allows the potential for system change in the most important industries to be assessed, while maintaining a high level of granularity and detail. "Ecosystem identification" delineates the change at the level of industry groups rather than individual companies, maximising the impact and ensuring the public character of innovation policy. "Platform-based smart and green transition" strengthens the ecosystem perspective with technologies and solutions over which many organisations can build complementary products and services.

Assessing the pathway in Greece and Cyprus, we showed its feasibility and functionality. Prioritisation worked as foreseen, enabling a focus on the most important industry groups; ecosystems and platforms for transition were identified within the priority industry groups. The ecosystem perspective is justified as the core component of the pathway, linking prioritisation and platform-based innovation and capitalising on the capacity of the digital transition to mobilise connected intelligence and capacity building in human–computer–community networks [82].

Industries and activity ecosystems in cities and regions are undergoing restructuring due to the widespread use of digital and green technologies, related products and processes, that can address contemporary challenges of growth, sustainability, and climate change. The pathway we described allows public authorities to assess the potential for smart and green transition at the level of each industry group without excluding any important group in advance. Two reasons justify the orientation of this approach: first, the widely accepted principle of smart specialisation for a place-specific innovation strategy or "one-size-does-not-fit-all", which suggests that theoretical predictions about future growth should be assessed and validated with place-specific data; second, the probability of finding innovative smart/green solutions in less expected activities, a trend outlined in many aspects of the innovation theory, such as the probabilistic and nondeterministic character of innovation, serendipity in innovation, and innovation outcomes by chaotic systemic combinations.

Funding: This research has received funding under a service contract with the European Commission (CCI Nr. 2018CE160AT103).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: I would like to thank George Peroulakis and the members of the Cyprus and Greece Unit of the Directorate General for Regional and Urban Policy of the European Commission (REGIO.G.3) for the collaboration and valuable comments on the two reports "Ecosystems and functioning EDP for S3 2021–2027" for Greece and Cyprus.

Conflicts of Interest: The author declares no conflict of interest.

References

1. European Commission. *The European Green Deal*; COM(2019) 640 Final; European Commission: Brussels, Belgium, 2019.
2. Pontikakis, D.; Fernandez, T.; Janssen, M.; Guy, K.; Marques Santos, A.; Boden, M.; Moncada-Paternò-Castello, P. *Projecting Opportunities for INdustrial Transitions (POINT): Concepts, Rationales and Methodological Guidelines for Territorial Reviews of Industrial Transition*; No JRC121439; Joint Research Centre: Ispra, Italy, 2020.
3. Frank, A.G.; Dalenogare, L.S.; Ayala, N.F. Industry 4.0 technologies: Implementation patterns in manufacturing companies. *Int. J. Prod. Econ.* **2019**, *210*, 15–26. [CrossRef]
4. Woyzbun, K.; Beitz, S.; Barnes, K. Industry transformation. In *Drivers of Change: For the Australian Labour Market to 2030: Proceedings of an Expert Scenario Forum*; The Academy of the Social Sciences in Australia: Canberra, Australia, 2014; pp. 17–35.
5. Oztemel, E.; Gursev, S. Literature review of Industry 4.0 and related technologies. *J. Intell. Manuf.* **2020**, *31*, 127–182. [CrossRef]
6. European Commission. *NACE Rev 2. Statistical Classification of Economic Activities in the European Community*; Office for Official Publications of the European Communities: Luxembourg, 2008; Available online: https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF (accessed on 26 June 2022).
7. Stephany, F. One size does not fit all: Constructing complementary digital reskilling strategies using online labour market data. *Big Data Soc.* **2021**, *8*, 20539517211003120. [CrossRef]
8. Oqubay, A.; Ohno, K. *How Nations Learn: Technological Learning, Industrial Policy, and Catch-Up*; Oxford University Press: Oxford, UK, 2019; p. 368.
9. Adner, R.; Kapoor, R. Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. *Strateg. Manag. J.* **2010**, *31*, 306–333. [CrossRef]
10. Tsujimoto, M.; Kajikawa, Y.; Tomita, J.; Matsumoto, Y. A review of the ecosystem concept—Towards coherent ecosystem design. *Technol. Forecast. Soc. Chang.* **2018**, *136*, 49–58. [CrossRef]
11. Gawer, A. The organization of technological platforms. In *Technology and Organization: Essays in Honour of Joan Woodward*; Phillips, N., Sewell, G., Griffiths, D., Eds.; Research in the Sociology of Organizations; Emerald Group Publishing Limited: Bingley, UK, 2010; Volume 29, pp. 287–296.
12. Castellani, S. Everything You Need to Know about Digital Platforms. Available online: http://stephane-castellani.com/everything-you-need-to-know-about-digital-platforms/ (accessed on 26 June 2022).
13. Gawer, A.; Cusumano, M.A. *Platform leadership: How Intel, Microsoft, and Cisco Drive Industry Innovation*; Harvard Business School Press: Boston, MA, USA, 2002; Volume 5, pp. 29–30.
14. Gawer, A.; Cusumano, M.A. Industry platforms and ecosystem innovation. *J. Prod. Innov. Manag.* **2014**, *31*, 417–433. [CrossRef]
15. Geels, F.W. Socio-technical transitions to sustainability: A review of criticisms and elaborations of the Multi-Level Perspective. *Curr. Opin. Environ. Sustain.* **2019**, *39*, 187–201. [CrossRef]
16. De Propris, L.; Bailey, D. Pathways of regional transformation and Industry 4.0. *Reg. Stud.* **2021**, *55*, 1617–1629. [CrossRef]
17. Bystrov, A.V.; Radaikin, A.G.; Fedoseev, E.V. Formation of organizational and economic model of cross-industry ecosystems. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2021; Volume 666, p. 062112.
18. D'Amico, G.; Arbolino, R.; Shi, L.; Yigitcanlar, T.; Ioppolo, G. Digitalisation driven urban metabolism circularity: A review and analysis of circular city initiatives. *Land Use Policy* **2022**, *112*, 105819. [CrossRef]
19. Yigitcanlar, T. *State of the Art and Future Perspectives in Smart and Sustainable Urban Development*; MDPI: Basel, Switzerland, 2022.
20. European Commission. *A New Industrial Strategy for Europe*; COM(2020) 102 Final; European Commission: Brussels, Belgium, 10 March 2020.
21. Landabaso, M. Guest editorial on research and innovation strategies for smart specialisation in Europe: Theory and practice of new innovation policy approaches. *Eur. J. Innov. Manag.* **2014**, *17*, 378–389. [CrossRef]
22. Bellandi, M.; De Propris, L. Local productive systems' transitions to industry 4.0. *Sustainability* **2021**, *13*, 13052. [CrossRef]
23. Fraga-Lamas, P.; Lopes, S.I.; Fernandez-Carames, T.M. Green IoT and edge AI as key technological enablers for a sustainable digital transition towards a smart circular economy: An industry 5.0 use case. *Sensors* **2021**, *21*, 5745. [CrossRef] [PubMed]

24. Komninos, N.; Panori, A.; Kakderi, C. Smart cities beyond algorithmic logic: Digital platforms, user engagement and data science. In *Smart Cities in the Post-algorithmic Era*; Edward Elgar Publishing: Cheltenham, UK, 2019.
25. Izzo, F.; Tomnyuk, V.; Lombardo, R. 4.0 digital transition and human capital: Evidence from the Italian fintech market. *Int. J. Manpow.* **2021**, *43*, 910–925. [CrossRef]
26. Favoretto, C.; Mendes, G.H.; Oliveira, M.G.; Cauchick-Miguel, P.A.; Coreynen, W. From servitization to digital servitization: How digitalization transforms companies' transition towards services. *Ind. Mark. Manag.* **2022**, *102*, 104–121. [CrossRef]
27. Kadir, B.A.; Broberg, O. Human well-being and system performance in the transition to industry 4.0. *Int. J. Ind. Ergon.* **2020**, *76*, 102936. [CrossRef]
28. Shamim, S.; Cang, S.; Yu, H.; Li, Y. Management approaches for Industry 4.0: A human resource management perspective. In Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC), Vancouver, BC, Canada, 24–29 July 2016; IEEE: New York, NY, USA, 2016; pp. 5309–5316.
29. Khan, A.; Turowski, K. A survey of current challenges in manufacturing industry and preparation for industry 4.0. In Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry (IITI'16), Sochi, Russia, 16–21 May 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 15–26.
30. Rojko, K.; Erman, N.; Jelovac, D. Impacts of the transformation to industry 4.0 in the manufacturing sector: The case of the US. Organizacija. *Home* **2020**, *53*, 287–305.
31. Martynov, V.V.; Shavaleeva, D.N.; Zaytseva, A.A. Information technology as the basis for transformation into a digital society and industry 5.0. In Proceedings of the 2019 International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS), Sochi, Russia, 23–27 September 2019; IEEE: New York, NY, USA, 2019; pp. 539–543.
32. Zhai, X.; An, Y. Analyzing influencing factors of green transformation in China's manufacturing industry under environmental regulation: A structural equation model. *J. Clean. Prod.* **2020**, *251*, 119760. [CrossRef]
33. Siddi, M. *The European Green Deal: Assessing Its Current State and Future Implementation*; FIIA Working Paper; FIIA: Helsinki, Finland, 2020.
34. Rivas, S.; Urraca, R.; Bertoldi, P.; Thiel, C. Towards the EU Green Deal: Local key factors to achieve ambitious 2030 climate targets. *J. Clean. Prod.* **2021**, *320*, 128878. [CrossRef]
35. Skjærseth, J.B. Towards a European Green Deal: The evolution of EU climate and energy policy mixes. *Int. Environ. Agreem. Politics Law Econ.* **2021**, *21*, 25–41. [CrossRef]
36. Cortright, J. New growth theory, new growth theory, technology and learning: Technology and learning. A practitioner's guide a practitioner's guide. *Rev. Econ. Dev. Lit. Pract.* **2001**, *1*, 1–40.
37. Acs, Z.; Sanders, M. Endogenous growth theory and regional extensions. In *Handbook of Regional Science*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 615–634.
38. OECD. *OECD Science, Technology and Innovation Outlook 2016*; OECD Publishing: Paris, France, 2016; Available online: https://doi.org/10.1787/sti_in_outlook-2016-en (accessed on 26 June 2022).
39. Geels, F.W. Technological transitions as evolutionary reconfiguration processes: A multi-level perspective and a case-study. *Res. Policy* **2002**, *31*, 1257–1274. [CrossRef]
40. Geels, F.W. *Transformative Innovation and Socio-Technical Transitions to Address Grand Challenges*; European Commission R&I Paper Series; Working Paper; European Commission: Brussels, Belgium, 2020.
41. Schot, J.; Geels, F.W. Strategic niche management and sustainable innovation journeys: Theory, findings, research agenda, and policy. *Technol. Anal. Strateg. Manag.* **2008**, *20*, 537–554. [CrossRef]
42. Weber, K.M.; Rohracher, H. Legitimizing research, technology and innovation policies for transformative change: Combining insights from innovation systems and multi-level perspective in a comprehensive 'failures' framework. *Res. Policy* **2012**, *41*, 1037–1047. [CrossRef]
43. Schot, J.; Steinmueller, W.E. Three frames for innovation policy: R&D, systems of innovation and transformative change. *Res. Policy* **2018**, *47*, 1554–1567.
44. Tukker, A. Leapfrogging into the future: Developing for sustainability. *Int. J. Innov. Sustain. Dev.* **2005**, *1*, 65–84. [CrossRef]
45. Watson, J.; Sauter, R. Sustainable innovation through leapfrogging: A review of the evidence. *Int. J. Technol. Glob.* **2011**, *5*, 170–189. [CrossRef]
46. Yu, Z.; Gibbs, D. Sustainability transitions and leapfrogging in latecomer cities: The development of solar thermal energy in Dezhou, China. *Reg. Stud.* **2018**, *52*, 68–79. [CrossRef]
47. Lim, S.B.; Yigitcanlar, T. Participatory Governance of Smart Cities: Insights from e-Participation of Putrajaya and Petaling Jaya, Malaysia. *Smart Cities* **2022**, *5*, 71–89. [CrossRef]
48. Iyer, A. Moving from Industry 2.0 to Industry 4.0: A case study from India on leapfrogging in smart manufacturing. *Procedia Manuf.* **2018**, *21*, 663–670. [CrossRef]
49. Primi, A.; Toselli, M. A global perspective on industry 4.0 and development: New gaps or opportunities to leapfrog? *J. Econ. Policy Reform* **2020**, *23*, 371–389. [CrossRef]
50. Komninos, N.; Kakderi, C.; Collado, A.; Papadaki, I.; Panori, A. Digital transformation of city ecosystems: Platforms shaping engagement and externalities across vertical markets. *J. Urban. Technol.* **2021**, *28*, 93–114. [CrossRef]
51. Komninos, N.; Kakderi, C.; Mora, L.; Panori, A.; Sefertzi, E. Towards high impact smart cities: A universal architecture based on connected intelligence spaces. *J. Knowl. Econ.* **2022**, *13*, 1169–1197. [CrossRef]

52. Das, A.; Dey, S. Global manufacturing value networks: Assessing the critical roles of platform ecosystems and industry 4.0. *J. Manuf. Technol. Manag.* **2021**, *32*, 1290–1311. [CrossRef]
53. Fuller, J.; Jacobides, M.G.; Reeves, M. The myths and realities of business ecosystems. *MIT Sloan Manag. Rev.* **2019**, *60*, 1–9.
54. Okano, M.T.; Antunes, S.N.; Fernandes, M.E. Digital transformation in the manufacturing industry under the optics of digital platforms and ecosystems. *Indep. J. Manag. Prod.* **2021**, *12*, 1139–1159. [CrossRef]
55. Voigt, K.I.; Müller, J.M. (Eds.) *Digital Business Models in Industrial Ecosystems: Lessons Learned from Industry 4.0 Across Europe*; Springer: Berlin/Heidelberg, Germany, 2021.
56. Andreoni, A.; Lazonick, W. Local Ecosystems and Social Conditions of Innovative Enterprise. In *The Oxford Handbook of Industrial Hubs and Economic Development*; Oxford University Press: Oxford, UK, 2020; pp. 77–97.
57. European Commission. *Updating the 2020 New Industrial Strategy: Building a Stronger Single Market for Europe's Recovery*; COM(2021) 350 Final; European Commission: Brussels, Belgium, 5 May 2021.
58. Rip, A.; Kemp, R. Technological change. *Hum. Choice Clim. Chang.* **1998**, *2*, 327–399.
59. Geels, F.W. *Technological Transitions and System Innovations: A Co-Evolutionary and Socio-Technical Analysis*; Edward Elgar Publishing: Cheltenham, UK, 2005.
60. Geels, F.W.; Schot, J. Typology of sociotechnical transition pathways. *Res. Policy* **2007**, *36*, 399–417. [CrossRef]
61. Genus, A.; Coles, A.M. Rethinking the multi-level perspective of technological transitions. *Res. Policy* **2008**, *37*, 1436–1445. [CrossRef]
62. Geels, F.W.; Kern, F.; Fuchs, G.; Hinderer, N.; Kungl, G.; Mylan, J.; Wassermann, S. The enactment of socio-technical transition pathways: A reformulated typology and a comparative multi-level analysis of the German and UK low-carbon electricity transitions 1990–2014. *Res. Policy* **2016**, *45*, 896–913. [CrossRef]
63. Geels, F.W.; Schot, J. Taxonomy of transition pathways in socio-technical transitions. In *Proceedings of the On Exploring Socio-Technical Transitions to Sustainability' Workshop*; Institute of Commonwealth Studies: London, UK, 2005.
64. Marinelli, E.; Fernández Sirera, T.; Pontikakis, D. *Towards a Transformative Smart Specialisation Strategy: Lessons from Catalonia, Bulgaria and Greece*; Publications Office of the European Union: Luxembourg, 2021.
65. Veldhuizen, C. Smart Specialisation as a transition management framework: Driving sustainability-focused regional innovation policy? *Res. Policy* **2020**, *49*, 103982. [CrossRef]
66. Berkhout, F.; Smith, A.; Stirling, A. Socio-technological regimes and transition contexts. *Syst. Innov. Transit. Sustain. Theory Evid. Policy* **2004**, *44*, 48–75.
67. Smith, A.; Stirling, A.; Berkhout, F. The governance of sustainable socio-technical transitions. *Res. Policy* **2005**, *34*, 1491–1510. [CrossRef]
68. Foray, D.; Goddard, J.; Goenaga Beldarrain, X.; Landabaso, M.; McCann, P.; Morgan, K.; Ortgea-Argiles, R. *Guide to Research and Innovation Strategies for Smart Specialisation (RIS 3), Smart Specialisation Platform*; IPTS Institute for Prospective Technological Studies, Joint Research Centre of the European Commission: Seville, Spain, 2012.
69. Komninos, N.; Tsampoulatidis, I.; Kakderi, C.; Nikolopoulos, S.; Kompatsiaris, I. Projects for smart cities: Ecosystems, connected intelligence and innovation for the radical transformation of cities. In *Smart Cities and Smart Communities*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 33–68.
70. Komninos, N.; Kakderi, C.; Panori, A.; Psaltoglou, A.; Chatziparadeisis, A. Ecosystems and functioning EDP for S3 2021–2027 in Greece. Report to the European Commission, DG Regional and Urban Policy. 2020. Available online: https://www.komninos.eu/wp-content/uploads/2022/07/ECOSYSTEMS-and-EDP-2021-2027-in-GREECE-v2020-05-16-Final.pdf (accessed on 24 June 2022).
71. Komninos, N. Ecosystems and functioning EDP for S3 2021–2027 in Cyprus. Report to the European Commission, DG Regional and Urban Policy. 2020. Available online: https://www.komninos.eu/wp-content/uploads/2022/07/ECOSYSTEMS-and-EDP-2021-2027-in-CYPRUS-v2020-05-16-FINAL.pdf (accessed on 24 June 2022).
72. Kakderi, C.; Komninos, N.; Panori, A.; Psaltoglou, A. Smart Specialisation 2.0: Driving public funds towards platforms and ecosystems. In Proceedings of the International Symposium: New Metropolitan Perspectives, Reggio Calabria, Italy, 18–23 May 2020; Springer: Cham, Switzerland, 2020; pp. 68–79.
73. Foray, D. From smart specialisation to smart specialisation policy. *Eur. J. Innov. Manag.* **2014**, *17*, 492–507. [CrossRef]
74. Reid, A.; Maroulis, N. From Strategy to Implementation: The Real Challenge for Smart Specialisation Policy. In *Advances in the Theory and Practice of Smart Specialisation*; Academic Press: Cambridge, MA, USA, 2017; pp. 293–318.
75. Komninos, N.; Kakderi, C.; Panori, A.; Garcia, E.; Fellnhofer, K.; Reid, A.; Cvijanović, V.; Roman, M.; Deakin, M.; Mora, L.; et al. Intelligence and co-creation in Smart Specialisation Strategies: Towards the next stage of RIS3. *archiDOCT* **2021**, *17*, 25361.
76. Panori, A.; Kakderi, C.; Dimitriadis, I. Combining technological relatedness and sectoral specialization for improving prioritization in Smart Specialisation. *Reg. Stud.* **2021**, 1–14. [CrossRef]
77. Boschma, R.; Frenken, K. Technological relatedness and regional branching. In *Beyond Territory: Dynamic Geographies of Knowledge Creation, Diffusion, and Innovation*; Bathelt, H., Feldman, M.P., Kogler, D.F., Eds.; Routledge: London, UK; New York, NY, USA, 2011; pp. 64–81.
78. Srnicek, N. *Platform Capitalism*; John Wiley and Sons: Hoboken, NJ, USA, 2017.
79. Komninos, N.; Tsarchopoulos, P.; Kakderi, C. New services design for smart cities: A planning roadmap for user-driven innovation. In Proceedings of the 2014 ACM International Workshop on Wireless and Mobile Technologies for Smart Cities, Philadelphia, PA, USA, 11 August 2014; pp. 29–38.

80. Lim, C.; Lee, J.H.; Sonthikorn, P.; Vongbunyong, S. Frugal innovation and leapfrogging innovation approach to the industry 4.0 challenge for a developing country. *Asian J. Technol. Innov.* **2021**, *29*, 87–108. [CrossRef]
81. El Bilali, H. The multi-level perspective in research on sustainability transitions in agriculture and food systems: A systematic review. *Agriculture* **2019**, *9*, 74. [CrossRef]
82. Komninos, N.; Panori, A. The creation of city smartness: Architectures of intelligence in smart cities and smart ecosystems. In *Smart Cities in the Post-Algorithmic Era*; Edward Elgar Publishing: Cheltenham, UK, 2019.

sustainability

MDPI

Article

Automatically Generating Scenarios from a Text Corpus: A Case Study on Electric Vehicles

Christopher W. H. Davis [1], Antonie J. Jetter [1] and Philippe J. Giabbanelli [2,*]

[1] Department of Engineering & Technology Management, Portland State University, Portland, OR 97201, USA; davis36@pdx.edu (C.W.H.D.); ajetter@pdx.edu (A.J.J.)
[2] Department of Computer Science and Software Engineering, Miami University, Oxford, OH 45056, USA
* Correspondence: giabbapj@miamioh.edu

Abstract: Creating 'what-if' scenarios to estimate possible futures is a key component of decision-making processes. However, this activity is labor intensive as it is primarily done manually by subject-matter experts who start by identifying relevant themes and their interconnections to build models, and then craft diverse and meaningful stories as scenarios to run on these models. Previous works have shown that text mining could automate the model-building aspect, for example, by using topic modeling to extract themes from a large corpus and employing variations of association rule mining to connect them in quantitative ways. In this paper, we propose to further automate the process of scenario generation by guiding pre-trained deep neural networks (i.e., BERT) through simulated conversations to extract a model from a corpus. Our case study on electric vehicles shows that our approach yields similar results to previous work while almost eliminating the need for manual involvement in model building, thus focusing human expertise on the final stage of crafting compelling scenarios. Specifically, by using the same corpus as a previous study on electric vehicles, we show that the model created here either performs similarly to the previous study when there is a consensus in the literature, or differs by highlighting important gaps on domains such as government deregulation.

Keywords: causal model; Fuzzy Cognitive Map; Q&A system; sustainability; text mining

Citation: Davis, C.W.H.; Jetter, A.J.; Giabbanelli, P.J. Automatically Generating Scenarios from a Text Corpus: A Case Study on Electric Vehicles. *Sustainability* **2022**, *14*, 7938. https://doi.org/10.3390/su14137938

Academic Editors: Tan Yigitcanlar, Bo Xia, Tatiana Tucunduva Philippi Cortese and Jamile Sabatini Marques

Received: 1 June 2022
Accepted: 27 June 2022
Published: 29 June 2022

1. Introduction

What-if questions are essential to making decisions by reasoning about the potential impacts of a situation. The situation could be an intervention (e.g., What happens to the sustainability of a city if we promote green spaces?) or a continuation of current trends (e.g., What happens in ten years if we continue with current emissions of pollutants?) [1]. A what-if question pertains to a specific *system*. For example, it would be impossible to answer the two questions above without a clear definition of the system (e.g., How do we measure sustainability? What is impacted by green spaces?). A *scenario* thus raises what-if questions of interest within the context of a clearly defined system, for example, by listing relevant factors and connecting them to track causal impacts. In other words, a scenario is a self-contained story about a potential future [2,3]. Scenarios have several demonstrated benefits for the decision-making activities of teams, such as raising awareness for the dynamics of an environment, managing uncertainty, evaluating different products, or breaking away from groupthink [4–7]. The field of *scenario planning* has articulated many approaches to craft such scenarios [8], often with the objective of producing a small number (typically 3–8) of plausible and alternative scenarios that cover different futures [9]. The quality of these scenarios is assessed through various criteria, such as plausibility [10], creativity [11], transparency [12], sufficient differentiation [13], relevance [14], or consistency [15].

A recurrent challenge is that scenario planning is a time-consuming and demanding process, for at least three reasons. First, the complexity of a system often calls for several subject-matter experts (SMEs), who are identified and involved via a trained facilitator to shed light on driving forces and current trends. Comprehensively understanding a system can thus be a significant endeavor, mobilizing numerous SMEs and necessitating the availability of a trained facilitator [16,17]. Second, there may be disagreements among SMEs on how some aspects of a system operate, or such mechanisms may simply by unknown. Similarly, some existing trends in the system or the actions planned by other stakeholders may not be known. There is thus a need to represent uncertainty. Third, under many scenario-planning techniques, teams focus on the 'big picture' to assess the futures of entire markets, industries, or even societies. While this is useful for high-level strategical thinking, it does not address the needs of teams who need more granular information to make tactical decisions related to specific products.

Given these challenges, there has been particular interest in automating some or all of the process of scenario planning, resulting in *Foresight Support Systems* [18,19]. Text collections have been an essential data source for such systems [20], as an indirect way to obtain vast amounts of domain expertise. This reflects a broader trend in future studies, which leverages unstructured data from websites, news posts, or academic journals [21–24]. These text collections have primarily been analyzed through web scrapping and topic modeling; recent examples include [25–28]. However, none of these studies *fully* automated the end-to-end process of scenario generation. For instance, [26] manually map the system, and [27,28] manually perform desk research and verification. Even works leveraging advances in natural language processing (NLP), such as BERT, contain a manual step of risk identification [29]. In this paper, we posit that there is a potential to go further in leveraging the information connected through massive text collection by using NLP to extract models of the system and craft scenarios.

In this paper, we improve the automatization of scenario generation by combining natural language processing and Fuzzy Cognitive Maps (FCMs). Our proposed tool is named SAAM, for Scenario Acceleration through Automated Modelling, and is available open source [30]. By emphasizing a fully automatic approach, we seek to drastically reduce the barriers to scenario development for teams who do not have the time or capacity to engage with subject-matter experts and trained facilitators.

To demonstrate the efficiency of our tool, we then apply it to a case study regarding electric vehicles (EVs). EVs were chosen as a guiding example for our technique as there is a demonstrated need and interest in scenario generation [31–33]. In particular, the scenarios covered by our case study include key themes about EVs, such as adoption [34–36], regulation and policy incentives [37–39], and technological enablers [40,41].

The remainder of this paper is structured as follows: To ensure that the manuscript is self-contained and usable both for computational scientists and sustainability specialists, our Background section provides the foundations for NLP and FCMs. Our Methods section builds on these foundations to introduce our proposed tool, SAAM. To demonstrate the efficiency of our tool, we then apply it to a case study regarding electric vehicles. Our results are compared with those obtained on the same corpus in a previous study performed by another group, showing that our model performs either similarly (with less manual involvement) or reveals important gaps. Our Discussion section contextualizes the potential of SAAM and outlines its limitations as well as opportunities for future improvements.

2. Background

2.1. Fuzzy Cognitive Maps

As evoked in the introduction, a scenario exists within the context of a clearly defined system. In other words, we need to *model* this system. Suitable modeling approaches fall into two broad categories. *Conceptual models* (e.g., causal maps, causal loop diagrams, mind maps) provide a structure to the system by identifying relevant factors and their interconnections [42–44]. Conceptual models have several benefits, such as identifying key factors in a system (e.g., via centrality), revealing themes (e.g., via community detection), or comparing perspectives (e.g., via Graph Edit Distance) [44–46]. However, these models offer limited support for scenario planning. For example, we can ask *what* will be impacted in a scenario, and we will follow links in the model to provide a list (e.g., via a Breadth-First Search). However, there is no quantification; hence, we cannot say whether some elements will be impacted more or less. In other words, the inability of a conceptual model to provide a quantitative estimate limits the decision-support tasks for which they are suitable. The second category of *quantitative (aggregate) models* offers these capabilities, but building them requires significantly more work [47]. Quantitative models are *simulation* models, which means that they can provide numerical answers by updating values based on certain rules. A well-known quantitative approach is System Dynamics [48], where the model runs differential equations to update concepts based on rates over time; this approach can provide highly accurate point-estimates, but requires significant quantitative data. Fuzzy Cognitive Maps (FCMs) do not include the notion of time; hence, they are simpler to build (e.g., entirely from qualitative data) at the expense of lower accuracy (i.e., cannot know exactly *when* an effect will be obtained) [49]. FCMs have been used in over 20,000 studies [50], including many works on scenario planning, as they provide quantitative system models that suffice to represent the driving forces that shape the future (e.g., technology, economy, social trends) and their interdependencies. Recent examples in sustainability include modeling the wind energy sector [51,52], social sustainability [53,54], planning viewed by rural communities [55] or urbanites [56], or managing waste flows [57]. Throughout these examples, the FCM is used for simulations by varying the input values to produce multiple scenarios; since the scenarios are all based on the same model, they are guaranteed to be internally consistent.

Mathematically, an FCM has two parts: a *causal structure* (similar to a conceptual model) and an *inference engine* (to run simulations). The causal structure is represented as a directed, weighted, labeled graph $G = (V, E)$, where V is the set of labeled nodes and E is the set of directed edges. Both nodes and edges have a weight. The weight of each node changes over each simulation step t to denote the extent to which a concept is present (1) or absent (1); it is denoted by $v_i^t \in [0, 1]$. The weight of each edge is held constant as it is considered a property of the system (e.g., if there are many anglers, then there are much less fish), whereas nodes correspond to a case (How many fish are there at a given point?). Edges are represented with an adjacency matrix, where $W_{i,j} \in [-1, 1]$ indicates the weight from node i to j. The weight is 0 if there is no relationship, positive if an increase in i causes an increase in j, and negative if an increase in i causes a decrease in j. The inference engine operates by synchronously updating all the nodes' values per Equation (1):

$$v_i^{t+1} = f\left(v_i^t + \sum_{j \in V, j \neq i} W_{j,i} \times v_j^t\right) \tag{1}$$

Intuitively, this update means that the next value of a node accounts for its current value (i.e., there is memory for one step), as well as the values of all incident nodes and the corresponding causal strengths. The function f serves to keep the output in the desired range $[0,\ 1]$. The update is performed until a stopping condition is met. The *desired* stopping condition is that a set of key nodes O (considered as outputs of the system) change by less than a user-defined value ε between two consecutive iterations. It is possible that this desired situation is not reached, due to oscillations or chaotic attractors. To ensure that the algorithm stops in any case, a secondary condition is a hard cap on the maximum number of iterations τ. Consequently, the updates stop if and only if Equation (2) holds true [58]:

$$\forall o \in O,\ \left| v_o^t - v_o^{t-1} \right| \leq \varepsilon \ or\ t \geq \tau \tag{2}$$

As the mathematics of FCMs have been abundantly covered elsewhere, we refer the reader to seminal reviews for further details [59,60]. In this paper, our interest is on (i) generating FCMs from text, and (ii) using them to craft scenarios. With regard to (i), we note that several works have extracted causal maps from text [26,61–63]; hence, they could generate the causal structure, but did not produce a complete FCM. Some works have focused on creating FCMs from summaries or large collection of documents [64,65], but they needed manual interventions (e.g., manual labeling, expert verification); hence, the process was only semi-automatic. The objective of (ii) building scenarios with FCMs is pursued by many studies [66–68], with several examining the role of FCMs as a communication tool to engage stakeholders in scenario generation [69,70].

2.2. Natural Language Processing

The major companies that own big data (e.g., Microsoft, Google, Amazon) have heavily invested in model creation and made several of the resulting models available to researchers and practitioners through their web services. For example, Google provides pre-trained models for natural language processing via its Natural Language AI. Pre-trained models in NLP often leverage deep neural networks, resulting in highly used models such as BERT or GPT [71,72]. BERT is of particular interest here, as it has previously been used to extract causal models from text [29]. We recently described BERT as follows [73]:

"BERT is a pre-trained deep bidirectional transformer, whose architecture consists of multiple encoders, each composed of two types of layers (multi-head self-attention layers, feed forward layers). To appreciate the number of parameters, consider that the text first goes through an embedding process (two to three dozen million parameters depending on the model), followed by transformers (each of which adds 7 or 12.5 million parameters depending on the model), ending with a pooling layer (0.5 or 1 million more parameters depending on the model). All of these parameters are trainable."

Intuitively, BERT models are trained by first creating a base model on a large unstructured dataset that can make predictions such as what word might appear next in a sentence. Secondly, the previous learnings are transferred, and models are fine-tuned on specific datasets that allow such functionality as answering questions based on the text in the dataset. To achieve this, BERT uses multiple layers of encoding so it can predict context and "understand" the difference between semantically similar terms such as "apple pie" or "apple tree" by encoding (1) the words, (2) the sentences, and (3) the positions of the words in the text. This combination of tokens is then fed into a neural network that creates the base model, which can be fine-tuned on specific text for NLP tasks. For a more detailed description of BERT, we refer the reader its highly cited source [74].

The core idea of repurposing BERT to extract a causal model is to build a question-answering (Q&A) system [75] in which we ask the question of what 'causes' or 'results' from a given factor, and then repeat the process on these causes and consequences to gradually build a model. In other words, a Q&A system can determine connections and causality between concepts in the model. By asking the system, "why do people buy more electric cars?" a human user identifies a concept of interest through the question—in this case, "electric cars". Q&A systems provide the answer by treating a pre-selected text corpus as the context. In this example, the corpus would focus on the electric car industry.

To briefly illustrate this notion within the context of sustainability, consider the following example of the fashion supply chain and the guiding question, "What causes pollution?" By applying a Q&A BERT-based model from the Hugging Face project [76] on online books about the fashion supply chain, we obtain a sample output such as in Table 1. Items in the 'answer' columns are concepts, the 'confidence' is the degree of certainty with which the algorithm identified the answer, and the 'context' provides an excerpt from the most relevant document containing the answer. In this example, "fast fashion brands" is returned with high confidence because it is directly referenced in the text as a cause of pollution, whereas very low confidence was returned for the other concepts because they are mentioned together but do not answer the question based on the text provided. The more text that associates fast fashion brands with pollution, the higher the confidence value would be. The context can also help to identify more relevant concepts, which can be used for further questions [77]. For instance, 'sustainable development' is mentioned as part of the answer 'global climate change', and it could lead to another line of questioning by asking the Q&A system, "What types of sustainable development are happening in the fashion industry?".

Table 1. Sample output from an NLP Q&A system when asked, "What causes pollution to increase?".

Answer	Confidence	Context
Fast fashion brands	0.489	on the other hand, fast fashion brands such as h & m, Zara, Topshop, have been blamed for creating poor labor welfare, severe environmental pollution as well as a massive amount of clothing disposal at the end of the product life cycle.
Global climate change	0.00713	introduction due to the aggravation of environmental pollution and global climate change, sustainable development has attracted more and more attention.
Overconsumption of energy	0.00669	by doing so, these companies alleviate conflicts of interest among participants and reduce pollution and overconsumption of energy.

3. Design of the Proposed SAAM tool

Overview

Our work seeks to automate the process of scenario generation. However, the analysts still need to be involved in defining the question and pointing to acceptable data sources. From that point onward, the automatic process can run. Overall, our proposed SAAM tool is composed of three stages: setup (which is manual), model building (which is automatized), and model use by humans (Figure 1); each of these stages is explained in a dedicated subsection below. Several parameters are involved in these stages, as summarized in Table 2. In short, the automation collects the data, runs the Q&A algorithms to find traceable answers from the text corpus, and builds the initial model as a Fuzzy Cognitive Map. People can inspect the answers, define filters, and potentially ask more questions to build out the model further. Once the model is fully built, people use it to run their scenarios. This process promotes an interplay of human interaction and Artificial Intelligence, hence following the human-in-the-loop approach that is increasingly promoted in machine learning to create more explainable models [78,79].

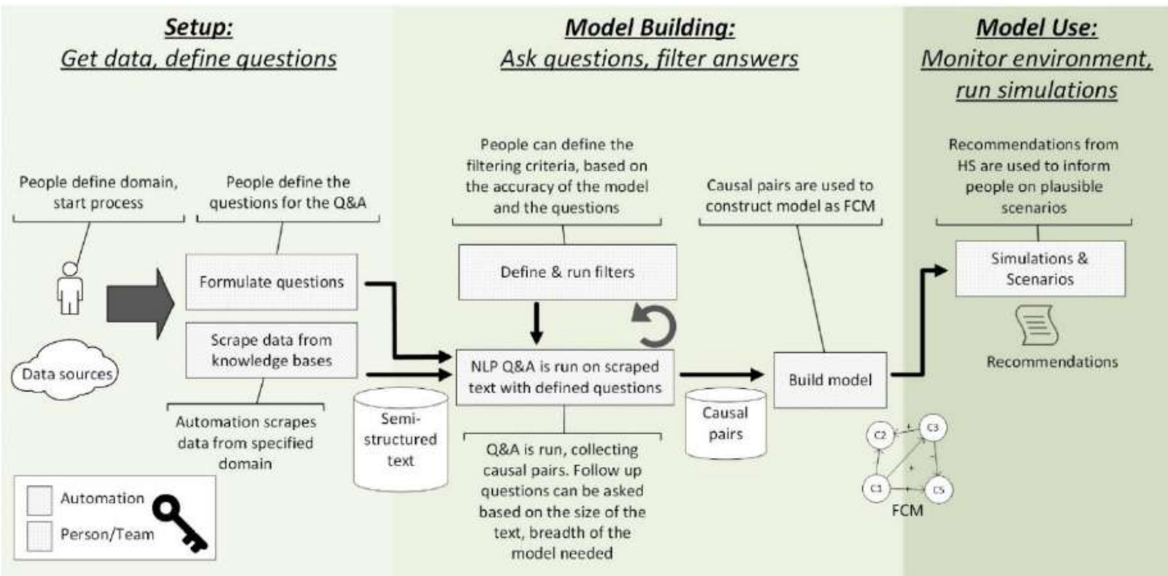


Figure 1. Overview of our proposed SAAM system.

Phase 1: Setup by defining questions and identifying relevant data sources.

Depending on the application domain, the modeling team starts by determining the questions to ask. This does not depend on their computational knowledge. It may depend on the stakeholders and commissioners, as is the case for any modeling endeavor [80]. For instance, if the modeling team seeks to better understand the future of self-driving vehicles, then they may ask questions that contain key terms such as "self-driving", "vehicles", or "self-driving cars". That is, they are responsible for identifying a set of seed concepts (or "nodes" of an FCM) belonging to the domain. If the modeling team is unsure about keywords that characterize a domain, they can also use NLP on relevant documents to extract candidate keywords, for instance, by removing stop-words and then extracting keywords with high frequency using libraries such as RAKE or Gensim. The keywords need to be structured into a question that can be passed onto a Q&A system. Two main options are as follows: If the team seeks a model to perform cause-and-effect analyses, then they may start with questions such as "what causes [phenomenon] to increase" and its complementary "what causes [phenomenon] to decrease"; this is similar to a facilitated modeling process investigating risks and protective factors [43]. Alternatively, if the team seeks a model that explores drivers for a specific technology, then they define questions based on the Political, Economic, Social, Technical, Environmental, and Legal (PESTEL) aspects of the technology. The PESTEL framework has been commonly used in scenario planning [81,82] and will be exemplified in our case study.

The modeling team also identifies appropriate data sources. These may include journal articles, newspaper articles, or websites that provide detailed information for the target domain.

Phase 2: Model building through the Q&A System and filtering.

The modeling team is responsible for specifying the number of iterations through which the system should build a model (i.e., 'question depth'). For example, after finding that A causes B, the model could be expanded to know what causes B, leading to another round of questions on increasing and decreasing causes of B; this would constitute a question depth of 1 (Figure 2). A modeler may choose a higher question depth if they only have a single question to start with, or if the corpus used is very large. After a certain number of iterations, answers typically start to decrease in confidence because they reach the knowledge limits of the corpus.

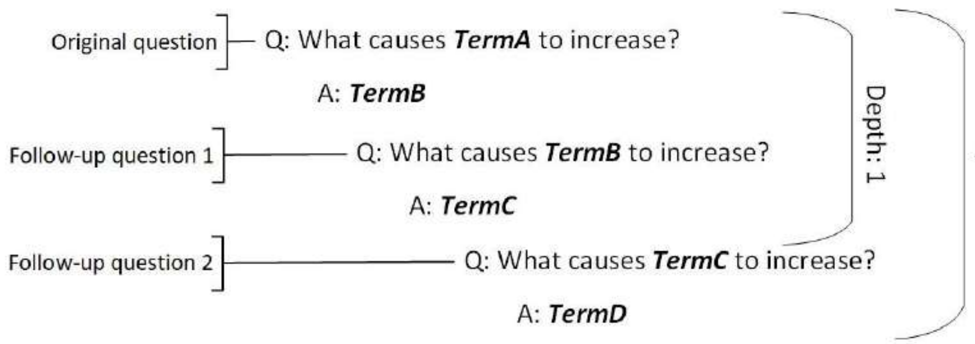


Figure 2. The 'question depth' is a key parameter governing model complexity.

Given (i) the corpus and (ii) the set of questions originating from phase 1, as well as (iii) the question depth, we use an NLP Q&A system to repeatedly find connections between concepts. Our work specifically uses the Hugging Face Q&A pipelines, but implementations can also be achieved via other open-source solutions such as Sentence Transformers [83]. When a factor X is identified as increasing Y, then we create an edge from X to Y with the value 1; conversely, if X decreases Y, then the edge has the value −1. Tracking the polarity of the relationship is important to later create the FCM.

Similar to the example in Table 1, the Q&A system responds to each question by providing the answer, together with a confidence level between 0 and 1, indicating the probability that the model got the correct answer, and token markers indicating where in the document the answer was found. For example, if a document contains the sentence "Pollution is a direct cause of a lower standard of living," and the Q&A algorithm asks the question, "what causes lower standards of living?", the model will return "pollution" as the answer, a high probability such as 0.89, and the beginning position in the document to where the answer was found. From these values, the answer and confidence score are directly relevant to assisting the modeler, and the token marker can be used to find the sentence and the document the answer was found in to give people using SAAM the full context of the answer. In this example, it is as if the model is saying "I am pretty sure that pollution is the answer because of this excerpt from the text you showed me". If responses were unfiltered, three problems could occur. First, answers with low confidence could be included, resulting in *noise* in the model. Second, words that look different but actually have the same meaning would be kept separately, hence resulting in a seemingly comprehensive but actually *redundant* model. Third, the name of a concept is usually a noun, but answers may consist of other types of words such as adjectives, which would be *harder to interpret* as labels in a causal model (e.g., the noun 'height' would be preferable to the adjective 'tall').

We handle these three situations through three filters, whose values can be set by the user. First, to avoid noise, the modeler may only keep connections that were returned with a high degree of confidence, thus filtering out results whose confidence is below a user-defined *confidence threshold*. The threshold depends on the Q&A model used and the corpus; hence, it should only be determined by the modeler after reviewing the initial results. Second, to avoid redundancy, the user provides a *semantic similarity threshold* between concepts such that answers above this value are deemed similar and merged. The semantic distance can be defined using Levenshtein or cosine distances. Our implementation uses the Levenshtein distance provided by the fuzzywuzzy library in Python [84], where a threshold of 100 is an exact match, and the closer to 0, the larger the distance between words. Finally, Part Of Speech (POS) tagging gives us the type for each word, and the user can *filter out POS* that do not belong to a causal model. We use the spaCy library [85] for this purpose. The default filter removes adjectives, punctuation, particles, symbols, and interjections.

Table 2. Parameters and inputs to our proposed SAAM system. Additional libraries were used for the system as a whole: Azure Machine Learning for data hosting and computing, and Machine Learning Pipelines for coordination of tasks.

Parameter/Input	Values	Purpose	Libraries Involved
Seed questions	String	The modeling team must define the problem of interest, which anchors the model.	N/A
Text collection	Resource set	Natural language processing is performed over a text corpus. It can be provided directly (e.g., as files or URLs) or retrieved from databases with search keywords.	Power Automate Desktop (to automate data collection)
Confidence threshold	[0, 1]	Filter results based on the confidence returned by the Q&A algorithm. The threshold range will vary based on the context; thus, the cut-off is up to the modeler.	Hugging Face Q&A

Table 2. *Cont.*

Parameter/Input	Values	Purpose	Libraries Involved
Semantic similarity threshold	[0, 100] (100 indicates perfect match)	Combine concepts that are semantically similar. This can use Levenshtein or cosine distance. A lower value will group many concepts, a higher value may create a model with different concepts but similar meanings.	fuzzywuzzy
POS Filtering	Array of POS tags (universal POS tags [86])	Parts of speech may be returned as answers, but would not make intuitive sense as concepts. In addition, aggregate models often limited to only using nouns as concepts.	spaCy

Phase 3: Using the model.

Phase 2 produces a model in the form of a Fuzzy Cognitive Map. As explained in our background, scenarios can be built using this FCM, based on situations that are currently considered by stakeholders. This is illustrated in the next section through our application of SAAM to electric vehicles.

4. Methods: Applying SAAM to Study Electric Vehicles

4.1. Overview

Our case study demonstrates the ability of our proposed SAAM system to extract concepts and causal links from a text, structure them into an FCM model, and use the model to run simulations on alternative future scenarios that are plausible, decision-relevant, and cover the range of uncertainty. For a fair comparison of the results obtained by SAAM, our case study follows the published work of another research team, such that we have matching objectives (study of electric vehicles), but different techniques. Specifically, the prior work used the PESTEL framework, followed by Latent Semantic Analysis (LSA) and Fuzzy Association Rule Mining to build a model *semi*-automatically [87]. The differences between their work and our approach are visually summarized in Figure 3. Most importantly, concept mapping was a manual endeavor in the previous study, while our work seeks to automatize this task as part of model building. Consequently, our comparison of SAAM's output with the previous study seeks to determine whether a more automatic approach can yield a similar model. Our workflow is summarized in Figure 4 and detailed in the following subsections.

4.2. System Setup: Data Sources, Seed Questions, Parameters

The authors of the comparison study did not publish the data they used. Consequently, we reconstructed the datasets from their description. Specifically, they scraped five websites: Siemens [88], MIT Technology Review [89], Kurzweil Accelerating Intelligence [90], World Future Society [91], and FutureTimeLine [92]. These sites were used by the authors of the prior work because they all provided articles that were future-oriented, hence, already containing an analysis of trends and expert insight on potential futures. Note that the prior work was published in 2016; hence, it would not be a fair comparison if we built a model based on the data available up to today (2022). In addition, some of the websites have ceased to exist, hence content may not only have expanded but also have been deleted. Consequently, we used the web archive Wayback Machine to re-create a dataset that most closely resembles the content available to authors of the prior work [91]. Specifically, we (i) only scraped articles discussing electric vehicles or alternative energy, as this filter was noted by the authors of the prior work; and (ii) we used the Wayback Machine to scrape data that would have been available as of March 2016.

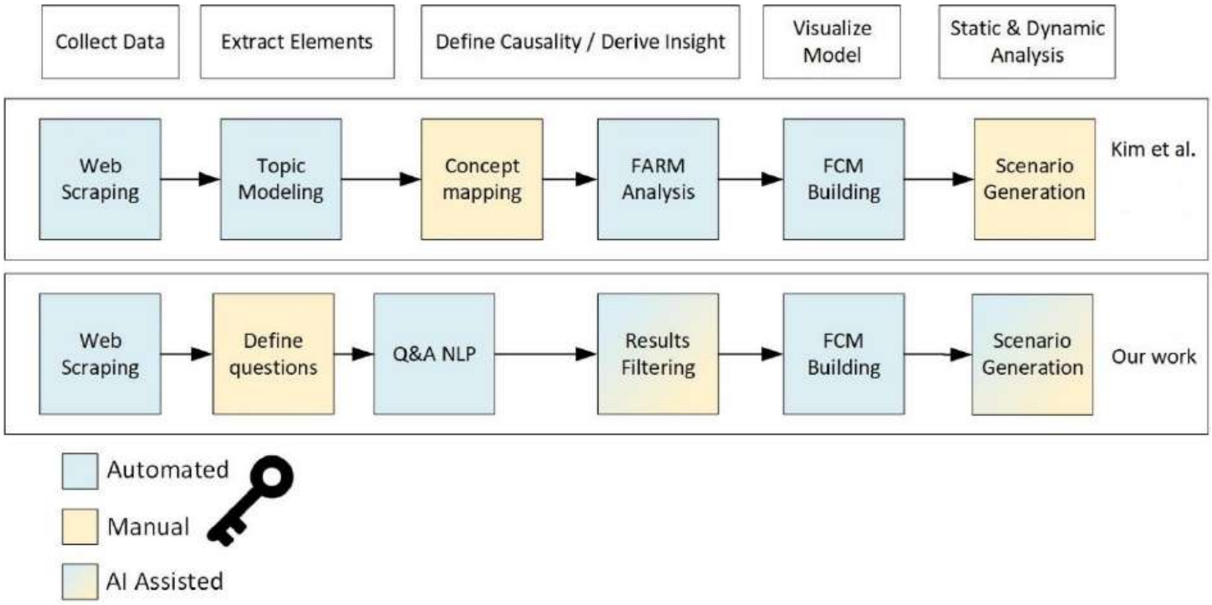


Figure 3. Comparison between our study (bottom row) and the prior work of Kim and colleagues [87].

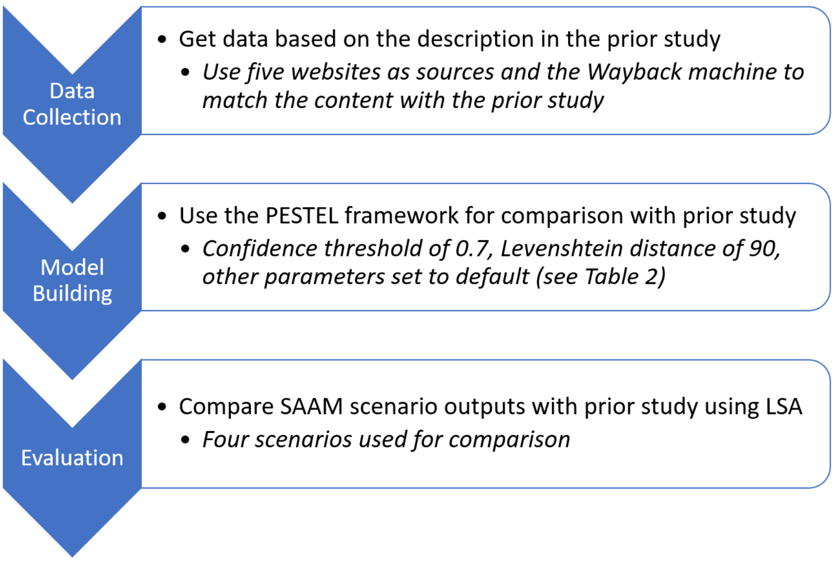


Figure 4. Key steps and specifications for our comparative study on electric vehicles.

Since the prior work used the PESTEL framework for its guiding questions, we also started by creating a set of questions about electric vehicles for each aspect of PESTEL. For example, under the environmental category, we asked, "What are benefits to the environment" and its complementary "What hurts the environment?" The full list of seed questions for our Q&A system is provided in Table 3. Parameter values for SAAM are listed in Figure 4.

Table 3. Seed questions based on PESTEL for our comparative study on electric vehicles.

Question	Subject	Weight	PESTEL Category
What technology is needed for electric vehicles?	EV adoption	1	Technology
Why use an electrified vehicle?		1	Open
What are impediments?		−1	
What are political factors?		1	Political
What are the benefits to the environment?		1	Environmental
What hurts the environment?	Environment	−1	
What are social benefits?	EV adoption	1	Social
What are social problems?		−1	
What social aspects affect electric vehicles?		1	
What are the economic benefits?	Economy	1	Economic
What are economic problems?		−1	
What are economic drivers?		−1	
What are legal problems?	EV adoption	−1	Legal
What are legal drivers?		1	
What are legal benefits?		1	

4.3. Comparison: Model Content and Simulated Scenarios

Models can be compared on the basis of their structure (e.g., which variables do they include? How are they connected?) and outputs (e.g., given the same input, which results do they produce?).

To compare the structure of the models, we examined the terms that they contained. To guide the comparison, we grouped the content of the SAAM model using the same categories as in the prior work. We stress that *our objective is not to find models with the same structure*. Rather, the structural comparison can tell us whether the models include similar categories, or aspects where one model was more comprehensive than the other. In contrast, we do *expect more similarities when comparing the output* of the models. For each scenario, we ran the SAAM model by creating inputs corresponding to the ones used in the original study, and then we compared the outputs of the two models. The original study had four high-level scenarios: (1) application of EV to tourism, (2) failure to develop battery technology, (3) failure of EV adoption in general, and (4) relaxation of government regulation. Changes were necessary in our comparative study, for two reasons. First, the prior work grouped the terms "economy", "consumer", "customer", "growth", and "tourism" in the tourism category by assuming that tourism is driven by consumers and is directly related to the economy. To avoid this narrow assumption, we broadened the scenario to study economic factors. Second, scenarios (1) and (3) are actually linked because (1) studies the effects of widespread EV adoption, whereas (3) examines the failure of widespread EV adoption. If we performed two scenarios on the same aspect, then that specific aspect of the model would artificially be counted twice. Consequently, we ran simulations on three scenarios: (i) economic factors affecting EV adoption, whether the economy is good or bad; (ii) what happens if battery technology does not develop; and (iii) what happens if the government decides to not help the EV industry at all by removing any incentives for EV and stopping any regulation efforts to increase adoption.

5. Results

5.1. Structural Comparison: Content of the Models

After filtering, the model produced by SAAM resulted in 52 unique concepts with 110 connections, as compared to the 15 concepts and 44 connections from the original study. The terms identified are shown in Box 1. As described in the previous section, we start our comparison by applying the categories from prior work to group the terms found by SAAM. The comparison is shown in Table 4. SAAM identified some of the same terms that were identified in the original study (green highlights), but also found concepts that were not detected in the prior work. For example, SAAM identified aspects such as consumer confidence, infrastructure investments needed, and natural resources required to build required batteries. This more comprehensive assessment can provide deeper insight into the data and hence support the creation of more robust models. Asking specific questions about social impacts led to answers such as 'thinking globally and acting locally', which was not in the LSA method. On the other hand, a few of the topics identified only make sense when knowing the context; for instance, 'your gas guzzler' refers to today's cars that run on gas, while 'aboriginal training' came from an Australian article about retraining individuals from underserved communities to work in new jobs created by the electric vehicles industry. Note neither the list of terms identified by SAAM nor those covered in the original study claim to address every facet of electric vehicles; rather, they extract information from a corpus focused on technology development. For example, emerging aspects such as electric mobility education [93] were absent from the corpus; hence, they are also absent in the list of terms.

5.2. Scenario Comparisons

Numerical results for each scenario are provided in the Table A1. Note that in the deregulation scenario, results are only indicative since the system oscillates instead of reaching stable values.

The original study showed that applying EV to tourism resulted in increased employment, a better economy, lower pollution levels, and improved energy efficiency. However, none of the data had articles about tourism; hence, the SAAM model did not directly cover tourism. After noting that the original study grouped tourism with economic benefits (see Section 4), we broadened the **scenario to the economy**. Specifically, we set the constructs 'employment', 'business development', 'current unit sales', 'economic activity', 'economic and safety benefits', and 'wealth' to high in one case (good economy) and to low in the other (poor economy). The SAAM model output a different result than the original study, noting that in a *good economy* 'no exhaust emissions' are adopted, but 'greenhouse gas emissions' increase and negatively affect 'the air'. In addition, we got richer results with SAAM, through some of the concepts that were not identified in the prior work; for instance, 'think globally act locally' decreases in a good economy, 'public investment' increases, and 'lack of infrastructure' decreases (meaning that the infrastructure will start to improve). In a good economy, 'EV adoption' decreases and 'your gas guzzler' (representing existing gas-powered vehicles) increases. In a *bad economy*, the inverse happens. Although this may seem counterintuitive at first, the transparency of the SAAM model lets us realize that, while several variables (technology, consumer confidence, battery technology) are high, the focus on sustainability decreases and volatility in gas prices decreases, which ultimately hurts the adoption of EVs. In short, this scenario implies that in a good economy, several technological aspects improve (EV infrastructure, battery technology, energy efficiency), but there is no strong *drive for consumers to adopt* EV technology.

Table 4. Comparison of concepts found by our SAAM system with the prior work's use of LSA. Categories are taken from the prior work to facilitate the alignment of the two models. Simple matches are shown in green, while noting that additional terms are equivalent within this context.

Category	SAAM Concepts	LSA Concepts
Air pollution	greenhouse gas emissions, no exhaust emissions, the air, your gas guzzler, energy pollution	Temperature, environment, pollution, atmosphere, carbon dioxide emission, greenhouse gas, CO_2, eco
Alternative energy technology	clean renewable energy sources, polarization systems	Renewable energy, diesel, biofuel, biomass, geothermal, petroleum, gasoline, hybrid, photovoltaic, solar energy
Battery technology	batteries, power and mileage limits, recharge speed	Lithium battery, ion battery, acid battery, storage, battery life, lightweight, BMS, lithium ion battery
Charging technology	a comprehensive charge station network, generic supercharging stations	Wireless power, charger, recharge, power transmission, charger
Costs reduction	EVs cost, the falling price of batteries, incentives	Cost reduction, incentive, support, maintenance cost
Economic revenue	business development, current unit sales, wealth, economic activity	Economy, growth, sales, investment, revenue, GDP, trade, import, export
Energy efficiency	energy efficiency	Energy efficiency, energy consumption, efficiency improvement, energy density, mileage
Government regulation	carbon pricing, cities conservation, governments, incentives, public investment, regulation	Regulation, incentive, policy, government, limitation, standard, tax reduction, policy
Industry-university collaboration	scholarships, aboriginal training	Company, startup, university, laboratory, investment, partnership, entrepreneur, grid
Job creation	employment	Job, worker, manufacturing, services, employment
Motor technology	electric motor	Engine, inverter, magnet, DC, AC, torque, capacity, motor
Usability	information technology	Automation, sensor, network connection, software, comfort, assistant, internet
Public transportation	Self-driving vehicles	Transportation, electric bus, driver, passenger
Safety	economic and safety benefits, self-driving vehicles	Safety, driverless, collision, vibration, pressure, security, stability, obstacle warning, monitoring
Other	thinking globally and acting locally, a completely carbon neutral transportation option, biomimicry, confidence, durability, environmentally conscious citizens	
Application to tourism		Consumer, customer, tourism, growth, economy

Box 1. List of terms identified by SAAM.

'EV', 'a completely carbon neutral transportation option', 'a comprehensive charge station network', 'aboriginal training', 'artificial intelligence', 'batteries', 'biomimicry', 'business development', 'carbon pricing', 'cities conservation', 'clean renewable energy sources', 'confidence', 'current unit sales', 'durability', 'economic activity', 'economic and safety benefits', 'electric motor', 'employment', 'energy efficiency', 'energy pollution', 'environmentally conscious citizens', 'evs cost', 'fear', 'gaps', 'generic super charging stations', 'governments', 'greenhouse gas emissions', 'harmony', 'incentives', 'information technology', 'infrastructure', 'lack of hydrogen infrastructure', 'liability', 'no exhaust emissions', 'oil and gas volatility', 'polarisation systems', 'potential roadblocks', 'power and mileage limits', 'public investment', 'rare earth metals', 'recharge speed', 'regulation', 'remote communities', 'save lives', 'scholarships', 'self-driving vehicles', 'significant technology improvements', 'sustainability', 'the air', 'the falling price of batteries', 'the power and mileage limits', 'thinking globally and acting locally', 'traffic congestion', 'transform mobility', 'wealth', 'your gas guzzler'

In the **scenario where battery technology fails to develop**, the original study concluded that there will be less job creation, less tourism, a poor economy, and an increase in pollution. To investigate this scenario, we set the corresponding variables in our model to low: 'batteries', 'lithium-air batteries', 'lithium-ion', 'lithium-ion batteries', 'recharge speed', 'power and mileage limits', and 'energy efficiency'. SAAM also found that 'employment' decreased, and terms associated with the economy ('economic activity', 'business development', 'current unit sales', 'wealth') all ended on low values. However, as in the previous scenario, SAAM had an inverse relationship between the economy and the environment; hence, it forecasted a decrease in 'greenhouse gas emissions' with an accompanying increase in the quality of 'the air'. In this scenario, EV adoption starts to improve even though the cost of EVs ('EVs cost') is driven up. Although battery technology fails to improve, an increased desire for sustainable solutions ('sustainability') and growing investment from the government ('public investment') help to offset the high cost of EVs.

Finally, in the **scenario of relaxing government regulations**, the prior work concluded a reduction in costs, an increase in safety, and an increase in energy efficiency. We simulated this scenario by setting all relevant concepts to low ('regulation', 'incentive', 'policy', 'government', 'limitation', 'standard', 'tax reduction'). Our simulation produced a limit cycle rather than a stable state. This indicates that if the government does nothing, then consumers would oscillate between EV adoption and rejection as the environment shifts from one preference to another based on competing factors. This sensitivity of our model to regulation suggests that it is a key concept in the adoption of EVs; hence, it deserves particular consideration when examining future strategies.

6. Discussion

6.1. Findings and Implications

Examining future scenarios is necessary to support decision-making activities [4–7]. These scenarios are created by teams and run on quantitative causal models, which forecast potential effects based on the evidence base. Creating a model is thus the cornerstone of scenario generation, yet it has long been a labor-intensive task [8,9]. Several works have brought automation to this process [18,19], in particular by deriving models from an evidence base consisting of a text corpus [25–28]. The recent work of Feblowitz and colleagues at IBM [29] is the closest to our approach in numerous regards: starting from a set seed of concepts (or 'risk forces'), it automatically fetches documents (multiple times daily via the Watson Discovery service) and uses a Q&A system powered by Hugging Face's Transformers to extract a model, noting when concepts can be deemed equivalent. A key limitation in previous works is that several steps continue to be performed by humans, as is the case in [29] where (meta)data on causal relationships is obtained via a crowd-sourced questionnaire, whereas we use the weights from the Fuzzy Cognitive Maps. In this paper, we proposed a step further in automation by only asking the modeling team to provide the initial guiding questions and the evidence base, and then creating a model. We demonstrated that the model could be used to investigate scenarios, by focusing on a case study in electric vehicles (EVs). EVs were chosen as a guiding example since (i) they have been the subject of several studies involving carefully crafted scenarios [31,94], and (ii) a previous study [87] with partial automation offered a direct comparison point with the model produced by our approach.

There are two key differences between our proposed approach (SAAM) and the prior study, which used less automation and involved Latent Semantic Analysis (LSA). First, LSA is used to find topics in a text collection and group terms together. Our system is not designed to perform such grouping, as we instead focus on finding terms by asking direct questions. The models are thus structured differently, with more granular content in SAAM offering a larger number of factors. However, it is possible that some of the content becomes too granular and needs to be interpreted given the context (e.g., 'the air'). Second, our proposed method and the previous one both have parameters that should be tuned by users. However, the methods are different; hence, the parameters offer control on different aspects. In SAAM, the modeling team can control filters, for instance, to force a simplification of the model by (i) combining semantically similar concepts and/or (ii) only accepting concepts where the system has high certainty. In contrast, the LSA method requires people to set a topic cluster size and manually name each final topic. Although our machine learning algorithm requires some human intervention to set parameters, we note that involving humans to train algorithms has been shown to facilitate co-learning between people and computers [95], and give analysts a better overall understanding of the model [96]. The potential benefits of a human-in-the-loop approach are noteworthy since our work is based on BERT, which is part of the set of artificial neural networks that have historically been characterized as 'lacking interpretability' and hence faced drawbacks in terms of trustworthiness by human decision makers [97].

Scenarios are supposed to help us step back and see the bigger picture, think outside the box, and consider alternatives that might not be obvious. Our results have shown that SAAM was able to generate alternative future scenarios that met this objective. We also demonstrated that the scenarios created via SAAM often agree with those created in the prior study, or propose a plausible line of reasoning when results differ. We emphasize that the application to electric vehicles provided a thorough evaluation of SAAM, but our tool is not limited to this specific application as it constitutes a reusable approach to generate scenarios. SAAM could thus be applied to similar issues in sustainability, such as autonomous vehicles [98], which have already been the subject of scenario generation studies using Fuzzy Cognitive Maps [99]. Our tool can more broadly benefit areas that frequently engage in the development of data-informed scenarios [100,101].

6.2. Limitations and Opportunities for Future Studies

One limitation of our comparison was the inability to use the same data as the original study, since it did not publish it. We re-created a dataset based on the sources and selection criteria mentioned, and ensured that it reflected what was available to the authors at the time. However, we did not detect any application to tourism in the evidence base; hence, this aspect was missing from the model and ultimately the scenario based on tourism was broadened to the economy.

The inspiring work by Feblowitz and colleagues suggests several improvements [29]. In particular, they were able to automatically generate trajectories from their model, using a planner and a clustering algorithm. To the best of our knowledge, planners able to generate a set of high-quality solutions (i.e., top-k planners) have not been applied to Fuzzy Cognitive Maps; hence, such algorithms would have to first be developed before we can produce trajectories.

The ability to transparently examine how the model reached a certain conclusion also holds particular promise for future studies. Indeed, the socio-environmental systems examined in sustainability studies are often complex, and models are at risk of becoming a 'black box' by being almost as complex. Maeda and colleagues stressed that "as the increasing complexity of models starts to influence policy making, it is important for scientists to create new approaches to communicate their underlying assumptions, reasoning, data and methods to stakeholders" [102]. Future work can thus contribute further to this communication component, for instance, by leveraging the Q&A system not only to build the model but also to ask how conclusions were reached.

7. Conclusions

Generating scenarios is essential for decision-making activities, but it involves a labor-intensive step of model building. We proposed a system (SAAM) that goes beyond previous automation initiatives, and we demonstrated that the system can result in well-formed scenarios by contrast to a previous study on electric vehicles. As the first manuscript detailing and applying SAAM, there are several opportunities for future work in improving components of the system or applying it for other fields of sustainability that heavily depend on scenario generations.

Author Contributions: C.W.H.D. developed the software and wrote the first draft. P.J.G. edited the manuscript. A.J.J. and P.J.G. supervised C.W.H.D. All authors approved the manuscript. P.J.G. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The code is available on our repository [30].

Acknowledgments: We thank Charles Weber and Ameeta Agrawal at Portland State University for their helpful feedback.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Each scenario is designed by setting the values of relevant factors in the model. For each scenario, we note the effect on other variables, as well as on the key construct of adopting electric vehicles (bottom row).

Categories	Concepts	Bad Economy	Good Economy	Battery Fail	Deregulation
air pollution	greenhouse gas emissions	−0.952398323	0.952398323	−0.957583063	−0.959324401
air pollution	no exhaust emissions	−0.952398323	0.952398323	−0.957583063	−0.959324401
air pollution	the air	0.957140415	−0.957140415	0.937957076	0.957801403
air pollution	your gas guzzler	−0.952398323	0.952398323	−0.957583063	−0.959324401
alternative energy technology	clean renewable energy sources	0.156727117	−0.156727117	0	0.121182442
alternative energy technology	polarisation systems	−0.952398323	0.952398323	−0.957583063	−0.959324401
battery technology	batteries	−0.691699732	0.691699732	−1	−0.646327877
battery technology	power and mileage limits	0.156727117	−0.156727117	−1	0.121182442
battery technology	recharge speed	0.156727117	−0.156727117	−1	0.121182442

Table A1. *Cont.*

Categories	Concepts	Bad Economy	Good Economy	Battery Fail	Deregulation
other	thinking globally and acting locally	0.156727117	−0.156727117	0	0.121182442
charging technology	a comprehensive charge station network	−0.952398323	0.952398323	−0.957583063	−0.959324401
charging technology	generic super charging stations	−0.691699732	0.691699732	0	−0.646327877
costs reduction	evs cost	0.957140415	−0.957140415	0.961179751	0.957801403
costs reduction	the falling price of batteries	0.957500995	−0.957500995	−0.929606356	0.932011183
economic activity	economic activity	−1	1	−0.985312975	0.990740486
economic revene	business development	−1	1	−0.957583063	−0.959324401
economic revene	current unit sales	−1	1	−0.957583063	−0.959324401
economic revene	wealth	−1	1	0	0.121182442
energy effeciency	energy efficiency	−0.952398323	0.952398323	−1	−0.959324401
energy pollution	energy pollution	0.388947408	−0.388947408	0.796604556	−0.510951584
government regulation	carbon pricing	0.156727117	−0.156727117	0	0.121182442
government regulation	cities conservation	0.873254834	−0.873254834	−0.774093871	−1
government regulation	governments	0.156727117	−0.156727117	0	−1
government regulation	incentives	−0.691699732	0.691699732	0	−1
government regulation	public investment	−0.902626096	0.902626096	0.686233755	−1
government regulation	regulation	−0.691699732	0.691699732	0	−1
industry-university collaboration	scholarships	−0.952398323	0.952398323	−0.957583063	−0.959324401
job creation	employment	−1	1	−0.90171281	0.940099166
motor technology	electric motor	0.972982612	−0.972982612	0.999909188	0.976648732
other	a completely carbon neutral transportation option	−0.952398323	0.952398323	−0.957583063	−0.959324401
industry-university collaboration	aboriginal training	−0.952398323	0.952398323	−0.957583063	−0.959324401
usability	artificial intelligence	0.156727117	−0.156727117	0	0.121182442
other	biomimicry	0.156727117	−0.156727117	0	0.121182442
other	confidence	−0.952398323	0.952398323	−0.957583063	−0.959324401
other	durability	0.156727117	−0.156727117	0	0.121182442
other	environmentally conscious citizens	0.156727117	−0.156727117	0	0.121182442

Table A1. *Cont.*

Categories	Concepts	Bad Economy	Good Economy	Battery Fail	Deregulation
usability	information technology	0.156727117	−0.156727117	0	0.121182442
other	infrastructure	0.96315824	−0.96315824	−0.774093871	0.995526376
other	lack of hydrogen infrastructure	0.924293982	−0.924293982	0	0.915954432
other	liability	0.156727117	−0.156727117	0	0.121182442
other	oil and gas volatility	0.156727117	−0.156727117	0	0.121182442
other	potential roadblocks	0.957140415	−0.957140415	0.937957076	0.957801403
other	rare earth metals	0.156727117	−0.156727117	0	0.121182442
other	remote communities	0.156727117	−0.156727117	0	0.121182442
other	significant technology improvements	−0.957479374	0.957479374	−0.060843278	−0.957345752
other	sustainability	0.255551223	−0.255551223	0.817909946	0.497931899
other	the power and mileage limits	0.924293982	−0.924293982	0.961179751	0.915954432
other	traffic congestion	0.156727117	−0.156727117	0	0.121182442
other	transform mobility	−0.952398323	0.952398323	−0.957583063	−0.959324401
public transportation	Self-driving vehicles	−0.74299687	0.74299687	0.542424672	−0.816448312
safety	economic and safety benefits	−1	1	−0.957583063	−0.959324401
	EV adoption	**0.901968281**	**−0.901968281**	**0.798453798**	**0.889187005**

References

1. Firmansyah, H.S.; Supangkat, S.H.; Arman, A.A.; Giabbanelli, P.J. Identifying the components and interrelationships of smart cities in Indonesia: Supporting policymaking via fuzzy cognitive systems. *IEEE Access* **2019**, *7*, 46136–46151. [CrossRef]
2. Bowman, G.; MacKay, R.B.; Masrani, S.; McKiernan, P. Storytelling and the scenario process: Understanding success and failure. *Technol. Forecast. Soc. Chang.* **2013**, *80*, 735–748. [CrossRef]
3. Van Notten, P. Scenario Development: A Typology of Approaches. In *Think Scenarios, Rethink Education*; OECD: Paris, France, 2006.
4. Derbyshire, J.; Giovannetti, E. Understanding the failure to understand New Product Development failures: Mitigating the uncertainty associated with innovating new products by combining scenario planning and forecasting. *Technol. Forecast. Soc. Chang.* **2017**, *125*, 334–344. [CrossRef]
5. Collier, Z.A.; Hendrickson, D.; Polmateer, T.L.; Lambert, J.H. Scenario analysis and PERT/CPM applied to strategic investment at an automated container port. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng.* **2018**, *4*, 04018026. [CrossRef]
6. Relich, M.; Bocewicz, G.; Rostek, K.; Banaszak, Z.A. A declarative approach to new product development project prototyping. *IEEE Intell. Syst.* **2020**, *36*, 88–95. [CrossRef]
7. Açikgöz, A.; Latham, G.P.; Acikgoz, F. Mediation of scenario planning on the reflection-performance relationship in new product development teams. *J. Bus. Ind. Mark.* **2020**, *36*, 256–268. [CrossRef]
8. Tiberius, V.; Siglow, C.; Sendra-García, J. Scenarios in business and management: The current stock and research opportunities. *J. Bus. Res.* **2020**, *121*, 235–242. [CrossRef]
9. Lindgren, M.; Bandhold, H. (Eds.) Scenario Planning in Practice. In *Scenario Planning: The Link between Future and Strategy*; Palgrave Macmillan: London, UK, 2009; pp. 49–117.
10. Spaniol, M.J.; Rowland, N.J. Defining scenario. *Futures Foresight Sci.* **2019**, *1*, e3. [CrossRef]
11. Alcamo, J.; Henrichs, T. Chapter two towards guidelines for environmental scenario analysis. *Dev. Integr. Environ. Assess.* **2008**, *2*, 13–35. [CrossRef]
12. Durance, P.; Godet, M. Scenario building: Uses and abuses. *Technol. Forecast. Soc. Chang.* **2010**, *77*, 1488–1492. [CrossRef]
13. Godet, M.; Roubelat, F.; Editors, G. Scenario Planning: An Open Future. *Technol. Forecast. Soc. Chang.* **2000**, *65*, 1–123. [CrossRef]
14. Van der Heijden, K. *Scenarios: The Art of Strategic Conversation*; John Wiley & Sons: Hoboken, NJ, USA, 2005.
15. Bradfield, R.; Derbyshire, J.; Wright, G. The critical role of history in scenario thinking: Augmenting causal analysis within the intuitive logics scenario development methodology. *Futures* **2016**, *77*, 56–66. [CrossRef]

16. Giabbanelli, P.J.; Galgoczy, M.C.; Nguyen, D.M.; Foy, R.; Rice, K.L.; Nataraj, N.; Brown, M.M.; Harper, C.R. Mapping the complexity of suicide by combining participatory modeling and network science. In Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Online, 8–11 November 2021; pp. 339–342.
17. Hedelin, B.; Gray, S.; Woehlke, S.; BenDor, T.; Singer, A.; Jordan, R.; Zellner, M.; Giabbanelli, P.; Glynn, P.; Jenni, K.; et al. What's left before participatory modeling can fully support real-world environmental planning processes: A case study review. *Environ. Model. Softw.* **2021**, *143*, 105073. [CrossRef]
18. Keller, J.; Markmann, C.; von der Gracht, H.A. Foresight support systems to facilitate regional innovations: A conceptualization case for a German logistics cluster. *Technol. Forecast. Soc. Chang.* **2015**, *97*, 15–28. [CrossRef]
19. Von der Gracht, H.A.; Bañuls, V.A.; Turoff, M.; Skulimowski, A.M.; Gordon, T.J. Foresight support systems: The future role of ICT for foresight. *Technol. Forecast. Soc. Chang.* **2015**, *97*, 1–6. [CrossRef]
20. Kayser, V.; Blind, K. Extending the knowledge base of foresight: The contribution of text mining. *Technol. Forecast. Soc. Chang.* **2017**, *116*, 208–215. [CrossRef]
21. Berg, S.; Wustmans, M.; Bröring, S. Identifying first signals of emerging dominance in a technological innovation system: A novel approach based on patents. *Technol. Forecast. Soc. Chang.* **2019**, *146*, 706–722. [CrossRef]
22. Jeong, Y.; Park, I.; Yoon, B. Identifying emerging Research and Business Development (R&BD) areas based on topic modeling and visualization with intellectual property right data. *Technol. Forecast. Soc. Chang.* **2019**, *146*, 655–672. [CrossRef]
23. Kose, T.; Sakata, I. Identifying technology convergence in the field of robotics research. *Technol. Forecast. Soc. Chang.* **2019**, *146*, 751–766. [CrossRef]
24. Robinson, D.K.; Lagnau, A.; Boon, W.P. Innovation pathways in additive manufacturing: Methods for tracing emerging and branching paths from rapid prototyping to alternative applications. *Technol. Forecast. Soc. Chang.* **2019**, *146*, 733–750. [CrossRef]
25. Ulman, M.; Šimek, P.; Masner, J.; Kogut, P.; Löytty, T.; Crehan, P.; Charvát, K.; Oliva, A.; Bergheim, S.R.; Kalaš, M.; et al. Towards Future Oriented Collaborative Policy Development for Rural Areas and People. *AGRIS-Line Pap. Econ. Inform.* **2020**, *12*, 111–124. [CrossRef]
26. Son, C.; Kim, J.; Kim, Y. Developing scenario-based technology roadmap in the big data era: An utilisation of fuzzy cognitive map and text mining techniques. *Technol. Anal. Strat. Manag.* **2020**, *32*, 272–291. [CrossRef]
27. Kayser, V.; Shala, E. Scenario development using web mining for outlining technology futures. *Technol. Forecast. Soc. Chang.* **2020**, *156*, 120086. [CrossRef]
28. Gokhberg, L.; Kuzminov, I.; Khabirova, E.; Thurner, T. Advanced text-mining for trend analysis of Russia's Extractive Industries. *Futures* **2020**, *115*, 102476. [CrossRef]
29. Feblowitz, M.; Hassanzadeh, O.; Katz, M.; Sohrabi, S.; Srinivas, K.; Udrea, O. IBM Scenario Planning Advisor: A Neuro-Symbolic ERM Solution. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 16032–16034. Available online: https://ojs.aaai.org/index.php/AAAI/article/view/18003 (accessed on 6 February 2022).
30. Davis, C.D.W. Scenario Acceleration through Automated Modelling (SAAM). Available online: https://github.com/cwhd/fuzzy-cognitive-horizon-scanning (accessed on 6 February 2022).
31. Blumberg, G.; Broll, R.; Weber, C. The impact of electric vehicles on the future European electricity system—A scenario analysis. *Energy Policy* **2022**, *161*, 112751. [CrossRef]
32. Kafaei, M.; Sedighizadeh, D.; Sedighizadeh, M.; Fini, A.S. An IGDT/Scenario based stochastic model for an energy hub considering hydrogen energy and electric vehicles: A case study of Qeshm Island, Iran. *Int. J. Electr. Power Energy Syst.* **2022**, *135*, 107477. [CrossRef]
33. Yu, Z.; Lu, F.; Zou, Y.; Yang, X. Quantifying energy flexibility of commuter plug-in electric vehicles within a residence–office coupling virtual microgrid. Part II: Case study setup for scenario and sensitivity analysis. *Energy Build.* **2022**, *254*, 111552. [CrossRef]
34. Singh, V.; Singh, V.; Vaibhav, S. A review and simple meta-analysis of factors influencing adoption of electric vehicles. *Transp. Res. Part D Transp. Environ.* **2020**, *86*, 102436. [CrossRef]
35. Song, R.; Potoglou, D. Are Existing Battery Electric Vehicles Adoption Studies Able to Inform Policy? A Review for Policymakers. *Sustainability* **2020**, *12*, 6494. [CrossRef]
36. Asif, U.; Schmidt, K. Fuel Cell Electric Vehicles (FCEV): Policy Advances to Enhance Commercial Success. *Sustainability* **2021**, *13*, 5149. [CrossRef]
37. Capuder, T.; Sprčić, D.M.; Zoričić, D.; Pandžić, H. Review of challenges and assessment of electric vehicles integration policy goals: Integrated risk analysis approach. *Int. J. Electr. Power Energy Syst.* **2020**, *119*, 105894. [CrossRef]
38. Whiston, M.M.; Azevedo, I.M.L.; Litster, S.; Samaras, C.; Whitefoot, K.S.; Whitacre, J.F. Expert elicitation on paths to advance fuel cell electric vehicles. *Energy Policy* **2022**, *160*, 112671. [CrossRef]
39. Baumgarte, F.; Kaiser, M.; Keller, R. Policy support measures for widespread expansion of fast charging infrastructure for electric vehicles. *Energy Policy* **2021**, *156*, 112372. [CrossRef]
40. Van Mierlo, J.; Berecibar, M.; El Baghdadi, M.; De Cauwer, C.; Messagie, M.; Coosemans, T.; Jacobs, V.A.; Hegazy, O. Beyond the State of the Art of Electric Vehicles: A Fact-Based Paper of the Current and Prospective Electric Vehicle Technologies. *World Electr. Veh. J.* **2021**, *12*, 20. [CrossRef]
41. Bhatti, G.; Mohan, H.; Singh, R.R. Towards the future of smart electric vehicles: Digital twin technology. *Renew. Sustain. Energy Rev.* **2021**, *141*, 110801. [CrossRef]

42. De Pinho, H. Generation of Systems Maps: Mapping Complex Systems and Population Health. In *Systems Science and Population Health*; Oxford University Press: Oxford, UK, 2017; pp. 61–67.
43. Giabbanelli, P.J.; Rice, K.L.; Galgoczy, M.C.; Nataraj, N.; Brown, M.M.; Harper, C.R.; Nguyen, M.D.; Foy, R. Pathways to suicide or collections of vicious cycles? Understanding the complexity of suicide through causal mapping. *Soc. Netw. Anal. Min.* **2022**, *12*, 60. [CrossRef]
44. McGlashan, J.; Johnstone, M.; Creighton, D.; De La Haye, K.; Allender, S. Quantifying a Systems Map: Network Analysis of a Childhood Obesity Causal Loop Diagram. *PLoS ONE* **2016**, *11*, e0165459. [CrossRef]
45. Giabbanelli, P.J.; Tawfik, A.A.; Gupta, V.K. Learning Analytics to Support Teachers' Assessment of Problem Solving: A Novel Application for Machine Learning and Graph Algorithms. In *Utilizing Learning Analytics to Support Study Success*; Springer: Cham, Switzerland, 2019; pp. 175–199. [CrossRef]
46. Anjum, M.; Voinov, A.; Taghikhah, F.; Pileggi, S.F. Discussoo: Towards an intelligent tool for multi-scale participatory modeling. *Environ. Model. Softw.* **2021**, *140*, 105044. [CrossRef]
47. Voinov, A.; Jenni, K.; Gray, S.; Kolagani, N.; Glynn, P.D.; Bommel, P.; Prell, C.; Zellner, M.; Paolisso, M.; Jordan, R.; et al. Tools and methods in participatory modeling: Selecting the right tool for the job. *Environ. Model. Softw.* **2018**, *109*, 232–255. [CrossRef]
48. Lofgren, E. System Dynamics Models. In *Systems Science and Population Health*; Oxford University Press: Oxford, UK, 2017; pp. 77–86.
49. Mkhitaryan, S.; Giabbanelli, P.J.; de Vries, N.K.; Crutzen, R. Dealing with complexity: How to use a hybrid approach to incorporate complexity in health behavior interventions. *Intell. Med.* **2020**, *3*, 100008. [CrossRef]
50. Kininmonth, S.; Gray, S.; Kok, K. Expert modelling. In *The Routledge Handbook of Research Methods for Social-Ecological Systems*; Taylor & Francis: Oxfordshire, UK, 2021; p. 231.
51. Amer, M.; Daim, T.U.; Jetter, A. Technology roadmap through fuzzy cognitive map-based scenarios: The case of wind energy sector of a developing country. *Technol. Anal. Strat. Manag.* **2016**, *28*, 131–155. [CrossRef]
52. Zare, S.G.; Alipour, M.; Hafezi, M.; Stewart, R.A.; Rahman, A. Examining wind energy deployment pathways in complex macro-economic and political settings using a fuzzy cognitive map-based method. *Energy* **2022**, *238*, 121673. [CrossRef]
53. Nasirzadeh, F.; Ghayoumian, M.; Khanzadi, M.; Cherati, M.R. Modelling the social dimension of sustainable development using fuzzy cognitive maps. *Int. J. Constr. Manag.* **2020**, *20*, 223–236. [CrossRef]
54. Averbuch, B.; Thorsøe, M.H.; Kjeldsen, C. Using fuzzy cognitive mapping and social capital to explain differences in sustainability perceptions between farmers in the northeast US and Denmark. *Agric. Hum. Values* **2022**, *39*, 435–453. [CrossRef]
55. Papageorgiou, K.; Singh, P.K.; Papageorgiou, E.; Chudasama, H.; Bochtis, D.; Stamoulis, G. Fuzzy Cognitive Map-Based Sustainable Socio-Economic Development Planning for Rural Communities. *Sustainability* **2019**, *12*, 305. [CrossRef]
56. Aminpour, P.; Gray, S.A.; Beck, M.W.; Furman, K.L.; Tsakiri, I.; Gittman, R.K.; Grabowski, J.H.; Helgeson, J.; Josephs, L.; Ruth, M.; et al. Urbanized knowledge syndrome—erosion of diversity and systems thinking in urbanites' mental models. *npj Urban Sustain.* **2022**, *2*, 11. [CrossRef]
57. Morone, P.; Yilan, G.; Imbert, E. Using fuzzy cognitive maps to identify better policy strategies to valorize organic waste flows: An Italian case study. *J. Clean. Prod.* **2021**, *319*, 128722. [CrossRef]
58. Mkhitaryan, S.; Giabbanelli, P.J. How modeling methods for fuzzy cognitive mapping can benefit from psychology research. In Proceedings of the 2021 Winter Simulation Conference (WSC), Phoenix, AZ, USA, 12–15 December 2021; pp. 1–2.
59. Papageorgiou, E.I.; Salmeron, J.L. A Review of Fuzzy Cognitive Maps Research During the Last Decade. *IEEE Trans. Fuzzy Syst.* **2012**, *21*, 66–79. [CrossRef]
60. Felix, G.; Nápoles, G.; Falcon, R.; Froelich, W.; Vanhoof, K.; Bello, R. A review on methods and software for fuzzy cognitive maps. *Artif. Intell. Rev.* **2019**, *52*, 1707–1737. [CrossRef]
61. Kwon, H.; Park, Y. Proactive development of emerging technology in a socially responsible manner: Data-driven problem solving process using latent semantic analysis. *J. Eng. Technol. Manag.* **2018**, *50*, 45–60. [CrossRef]
62. Sandhu, M.; Giabbanelli, P.J.; Mago, V.K. From Social Media to Expert Reports: The Impact of Source Selection on Automatically Validating Complex Conceptual Models of Obesity. In *Society Computing and Social Media*; Meiselwitz, G., Ed.; Springer: Cham, Switzerland, 2019; Volume 11578.
63. Villalon, J.; Calvo, R.A. Concept maps as cognitive visualizations of writing assignments. *J. Educ. Technol. Soc.* **2011**, *14*, 16–27. Available online: https://www.jstor.org/stable/jeductechsoci.14.3.16?seq=1 (accessed on 6 February 2022).
64. Hajek, P.; Prochazka, O.; Pachura, P. Fuzzy cognitive maps based on text analysis for supporting strategic planning. In Proceedings of the 2017 International Conference on Research and Innovation in Information Systems (ICRIIS), Langkawi, Malaysia, 16–17 July 2017; pp. 1–6.
65. Pillutla, V.S.; Giabbanelli, P.J. Iterative generation of insight from text collections through mutually reinforcing visualizations and fuzzy cognitive maps. *Appl. Soft Comput.* **2019**, *76*, 459–472. [CrossRef]
66. Lavin, E.A.; Giabbanelli, P.J.; Stefanik, A.T.; Gray, S.A.; Arlinghaus, R. Should we simulate mental models to assess whether they agree? In Proceedings of the Annual Simulation Symposium, San Diego, CA, USA, 15–18 April 2018; pp. 1–12.
67. Kok, K. The potential of Fuzzy Cognitive Maps for semi-quantitative scenario development, with an example from Brazil. *Glob. Environ. Chang.* **2009**, *19*, 122–133. [CrossRef]
68. Jetter, A.; Schweinfort, W. Building scenarios with Fuzzy Cognitive Maps: An exploratory study of solar energy. *Futures* **2011**, *43*, 52–66. [CrossRef]

69. Van Vliet, M.; Kok, K.; Veldkamp, T. Linking stakeholders and modellers in scenario studies: The use of Fuzzy Cognitive Maps as a communication and learning tool. *Futures* **2010**, *42*, 1–14. [CrossRef]
70. Jetter, A.J.; Kok, K. Fuzzy Cognitive Maps for futures studies—A methodological assessment of concepts and methods. *Futures* **2014**, *61*, 45–57. [CrossRef]
71. Acheampong, F.A.; Nunoo-Mensah, H.; Chen, W. Transformer models for text-based emotion detection: A review of BERT-based approaches. *Artif. Intell. Rev.* **2021**, *54*, 5789–5829. [CrossRef]
72. Dale, R. GPT-3: What's it good for? *Nat. Lang. Eng.* **2021**, *27*, 113–118. [CrossRef]
73. Galgoczy, M.C.; Phatak, A.; Vinson, D.; Mago, V.K.; Giabbanelli, P.J. (Re)shaping online narratives: When bots promote the message of President Trump during his first impeachment. *PeerJ Comput. Sci.* **2022**, *8*, e947. [CrossRef]
74. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
75. Sabharwal, N.; Agrawal, A. BERT Model Applications: Question Answering System. In *Hands-on Question Answering Systems with BERT*; Apress: Berkeley, CA, USA, 2021.
76. Hugging Face Project. Available online: https://huggingface.co/models (accessed on 6 February 2022).
77. Sandhu, M.; Vinson, C.D.; Mago, V.K.; Giabbanelli, P.J. From associations to sarcasm: Mining the shift of opinions regarding the Supreme Court on twitter. *Online Soc. Netw. Media* **2019**, *14*, 100054. [CrossRef]
78. Xin, D.; Ma, L.; Liu, J.; Macke, S.; Song, S.; Parameswaran, A. Accelerating human-in-the-loop machine learning: Challenges and opportunities. In Proceedings of the Second Workshop on Data Management for End-to-End Machine Learning, Houston, TX, USA, 15 June 2018; pp. 1–4.
79. Wu, X.; Xiao, L.; Sun, Y.; Zhang, J.; Ma, T.; He, L. A survey of human-in-the-loop for machine learning. *Future Gener. Comput. Syst.* **2022**, *135*, 364–381. [CrossRef]
80. Treasury, H.M. *The Aqua Book: Guidance on Producing Quality Analysis for Government*; H. M. Government: London, UK, 2015.
81. Oliver, J.J.; Parrett, E. Managing future uncertainty: Reevaluating the role of scenario planning. *Bus. Horiz.* **2018**, *61*, 339–352. [CrossRef]
82. Norouzi, N.; Fani, M.; Ziarani, Z.K. The fall of oil Age: A scenario planning approach over the last peak oil of human history by 2040. *J. Pet. Sci. Eng.* **2020**, *188*, 106827. [CrossRef]
83. Reimer, N. SentenceTransformers. Available online: www.sbert.net (accessed on 6 February 2022).
84. SeatGeek. Fuzzy String Matching in Python. Available online: https://github.com/seatgeek/thefuzz (accessed on 6 February 2022).
85. Explosion. spaCy: Industrial-Strength Natural Language Processing. Available online: https://spacy.io/ (accessed on 6 February 2022).
86. Universal Dependencies Contributors. Universal POS Tags. Available online: https://universaldependencies.org/u/pos/ (accessed on 6 February 2022).
87. Kim, J.; Han, M.; Lee, Y.; Park, Y. Futuristic data-driven scenario building: Incorporating text mining and fuzzy association rule mining into fuzzy cognitive map. *Expert Syst. Appl.* **2016**, *57*, 311–323. [CrossRef]
88. Siemens. Company, Innovation. Available online: https://new.siemens.com/global/en/company/innovation.html (accessed on 6 February 2022).
89. MIT Technology Review. Available online: https://www.technologyreview.com/all-topics (accessed on 6 February 2022).
90. Kurzweil. Tracking the Acceleration of Intelligence. Available online: http://www.kurzweilai.net (accessed on 6 February 2022).
91. World Future Society. Available online: http://web.archive.org/web/*/http://www.wfs.org (accessed on 6 February 2022).
92. FutureTimeLine. Available online: http://futuretimeline.net/index.htm (accessed on 6 February 2022).
93. Turoń, K.; Kubik, A.; Chen, F. When, What and How to Teach about Electric Mobility? An Innovative Teaching Concept for All Stages of Education: Lessons from Poland. *Energies* **2021**, *14*, 6440. [CrossRef]
94. Seljom, P.; Kvalbein, L.; Hellemo, L.; Kaut, M.; Ortiz, M.M. Stochastic modelling of variable renewables in long-term energy models: Dataset, scenario generation & quality of results. *Energy* **2021**, *236*, 121415. [CrossRef]
95. Fiebrink, R.; Cook, P.R.; Trueman, D. Human model evaluation in interactive supervised learning. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada; 2011; pp. 147–156.
96. Robert, S.; Büttner, S.; Röcker, C.; Holzinger, A. Reasoning Under Uncertainty: Towards Collaborative Interactive Machine Learning. In *Machine Learning for Health Informatics. Lecture Notes in Computer Science*; Holzinger, A., Ed.; Springer: Cham, Switzerland, 2016; Volume 9605.
97. Merritt, S.M. Affective Processes in Human—Automation Interactions. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2011**, *53*, 356–370. [CrossRef]
98. Czech, P. Autonomous vehicles: Basic issues. *Sci. J. Silesian Univ. Technol. Ser. Transp.* **2018**, *100*, 15–22. [CrossRef]
99. Zhang, P.; Jetter, A. A framework for building integrative scenarios of autonomous vehicle technology application and impacts, using fuzzy cognitive maps (FCM). In Proceedings of the 2018 Portland International Conference on Management of Engineering and Technology (PICMET), Honolulu, HI, USA, 19–23 August 2018; pp. 1–14.
100. Chapaloglou, S.; Varagnolo, D.; Marra, F.; Tedeschi, E. Data-informed scenario generation for statistically stable energy storage sizing in isolated power systems. *J. Energy Storage* **2022**, *51*, 104311. [CrossRef]

101. Kamari, A.; Kirkegaard, P.H.; Schultz, C.P.L. PARADIS-A process integrating tool for rapid generation and evaluation of holistic renovation scenarios. *J. Build. Eng.* **2020**, *34*, 101944. [CrossRef]
102. Maeda, E.E.; Haapasaari, P.; Helle, I.; Lehikoinen, A.; Voinov, A.; Kuikka, S. Black Boxes and the Role of Modeling in Environmental Policy Making. *Front. Environ. Sci.* **2021**, *9*, 63. [CrossRef]

sustainability

MDPI

Article

Musawah: A Data-Driven AI Approach and Tool to Co-Create Healthcare Services with a Case Study on Cancer Disease in Saudi Arabia

Nala Alahmari [1], Sarah Alswedani [1], Ahmed Alzahrani [1], Iyad Katib [1], Aiiad Albeshri [1] and Rashid Mehmood [2,*]

[1] Department of Computer Science, FCIT, King Abdulaziz University, Jeddah 21589, Saudi Arabia; nsaadalahmari0001@stu.kau.edu.sa (N.A.); sabdullahalswedani@stu.kau.edu.sa (S.A.); asalzahrani@kau.edu.sa (A.A.); iakatib@kau.edu.sa (I.K.); AAAlbeshri@kau.edu.sa (A.A.)

[2] High-Performance Computing Center, King Abdulaziz University, Jeddah 21589, Saudi Arabia

* Correspondence: RMehmood@kau.edu.sa; Tel.: +966-537069318

Abstract: The sustainability of human existence is in dire danger and this threat applies to our environment, societies, and economies. Smartization of cities and societies has the potential to unite individuals and nations towards sustainability as it requires engaging with our environments, analyzing them, and making sustainable decisions regulated by triple bottom line (TBL). Poor healthcare systems affect individuals, societies, the planet, and economies. This paper proposes a data-driven artificial intelligence (AI) based approach called Musawah to automatically discover healthcare services that can be developed or co-created by various stakeholders using social media analysis. The case study focuses on cancer disease in Saudi Arabia using Twitter data in the Arabic language. Specifically, we discover 17 services using machine learning from Twitter data using the Latent Dirichlet Allocation algorithm (LDA) and group them into five macro-services, namely, Prevention, Treatment, Psychological Support, Socioeconomic Sustainability, and Information Availability. Subsequently, we show the possibility of finding additional services by employing a topical search over the dataset and have discovered 42 additional services. We developed a software tool from scratch for this work that implements a complete machine learning pipeline using a dataset containing over 1.35 million tweets we curated during September–November 2021. Open service and value healthcare systems based on freely available information can revolutionize healthcare in manners similar to the open-source revolution by using information made available by the public, the government, third and fourth sectors, or others, allowing new forms of preventions, cures, treatments, and support structures.

Keywords: machine learning; big data analytics; social media; Twitter; smart healthcare; cancer; Arabic language; Latent Dirichlet Allocation (LDA); topic modeling; Natural Language Processing (NLP); smart cities

Citation: Alahmari, N.; Alswedani, S.; Alzahrani, A.; Katib, I.; Albeshri, A.; Mehmood, R. Musawah: A Data-Driven AI Approach and Tool to Co-Create Healthcare Services with a Case Study on Cancer Disease in Saudi Arabia. *Sustainability* **2022**, *14*, 3313. https://doi.org/10.3390/su14063313

Academic Editor: Petra Poulová

Received: 22 February 2022
Accepted: 5 March 2022
Published: 11 March 2022

1. Introduction

The sustainability of human existence—our existence—is in dire danger and this threat applies to our environment, societies, and economies. The threats to the environment are evident in the deteriorating planet's conditions. The threats to our societies are apparent in the deteriorating physical and psychological health of individuals and groups; increasing frequency and magnitudes of conflicts, riots, and wars, the rise of materialism, and the deteriorating love, harmony, and sincerity between people. The risks to our economies are manifested in the changing foci from the net benefits that the wealth brings to society, to mere numbers counting products, earnings, spending, sales, exports, and GDPs. Nationalisms, racisms, gender, and other wars are on the rise and seen as important in the name of equality, freedom, and missions, and little thought is given that any group is comprised of individuals, and each individual is a fellow human who is juggling with many difficulties like anyone else.

If done right, an umbrella term that has the potential to unite individuals towards sustainability is smartization of cities, and societies that involve the transformation of our traditional environments into smarter ones [1,2]. Smartness is defined by its core objective of the triple bottom line (TBL)—i.e., social, environmental, and economic sustainability [3] and can be achieved by engaging with our environments, analyzing them, and making sustainable decisions regulated by triple bottom line [4,5]. Note that the terms quadruple bottom line and quintuple bottom line have also been used in the literature with additional emphasis on ethics, equity, and purpose (soul, spirituality, or culture). For simplicity, we take the view that equity, efficiencies, ethics, purpose, innovation, etc. are included in TBL, the three dimensions of sustainability, however, we understand that placing in various statements an emphasis on specific aspects such as equity and ethics can be beneficial. The requirements for and the definitions of smartness have evolved over time with a focus in the early days on the digital aspects of the systems, shifting gradually to incorporate efficiency, equity, and triple bottom line in it [6].

Human health and healthcare are the cornerstones of all three facets of sustainability, the TBL [7]. Poor health and healthcare systems affect individuals, societies, the planet, and economies [8,9]. Healthcare systems are under increasing pressure to reduce their social, economic, and environmental costs [7,8,10]. An increasingly substantial share of GDPs is being spent on healthcare by countries around the world [4]. The US spends around 20% of its economy on healthcare [11]. Comparing different industries, the healthcare sector is notoriously inefficient and wasteful of resources and budgets [11]. Both the mental and physical health of people around the world is declining with the surge in lifelong illnesses with ageing populations. The cost of healthcare is rising while the quality is declining. The COVID-19 pandemic has hindered many patients with chronic and terminal illnesses from getting proper treatment due to pandemic regulations [12]. Disparities and inequities in healthcare around the world are rising [13]. Most important of all, healthcare systems are a major contributor to the deterioration of our planet's environment due to high energy usage, disposable supplies, etc. This environmental damage causes health problems that need to be medically treated, while these treatments in turn further damage the environment, and thus a chain reaction [14].

There is a growing demand for the prevention of diseases and reducing the need for disease monitoring, screening, and treatment to a minimum [7–10]. This in turn requires an open space for innovations and dynamic interactions between healthcare stakeholders to understand and provide solutions, services, information and resource supply chains, and community support structures for timely interventions and disease prevention and progression. Fortunately, social media, which hosts around 60% of the world's population [15], is one such platform that provides an open space for stakeholder interactions.

Motivated by the urgency and gravity of the challenges facing healthcare, and the technological opportunities, this paper proposes a data-driven artificial intelligence (AI) based approach called Musawah (Musawah is an Arabic word meaning Equity. We call our approach equity as we believe that sustainability and sustainable healthcare can be achieved by people believing in and practicing equity. We believe that our approach of open service and value healthcare systems based on freely available information can enable equity and sustainability) to automatically detect and identify healthcare services that can be developed or co-created by various stakeholders using social media analysis. Essentially, the aim herein is to investigate the role of big data analytics over social media to automatically detect needs and value propositions that could be nurtured and turned into service co-creation processes through engaged participation over social and digital media, eventually co-creating services. The co-created services are based on values that are not necessarily materialistic, but are driven by equity, altruism, community strengthening, innovation, and social cohesion. The case study focuses on cancer disease in Saudi Arabia using Twitter data analytics in the Arabic language; however, the proposed approach broadly can be used for any disease or purpose and in any language.

Specifically, we detect 17 services (e.g., screening, hope and optimism, financial support, and awareness campaigns) using unsupervised machine learning from Twitter data using the Latent Dirichlet Allocation algorithm (LDA) and group them into five macro-services namely, Prevention, Treatment, Psychological Support, Socioeconomic Sustainability, and Information Availability. The Prevention macro-service includes five services—Early Diagnosis, Prevention and Control, Causes, Screening, and Symptoms—and involves measures that may avoid cancer (e.g., maintaining a healthy lifestyle and avoiding cancer-causing substances) or help recovery from cancer or slow down its progression (e.g., symptoms and early detection).

The second macro-service is Treatment and relates to various treatment options available for cancer. It includes two detected services: Chemo and Radiation Therapy and Surgical Therapy. It does not mean that only two treatment options or services are available. It simply shows that these two services were detected by our tool, which may indicate the two treatment services that are more common based on the dataset or its temporal dimensions, or it may indicate the need for new treatment services. The third macro-service is Psychological Support, which captures the services to support patients and their families to help them cope with their psychological needs, such as emotions and spirituality. It includes three services, Spiritual Support, Suffering, and Hope and Optimism. The fourth macro-service, Socioeconomic Sustainability, includes four services, namely, Government Support, Socioeconomic and Operational Challenges, Charity Organizations, and Financial Support. It relates to the financial, healthcare, and other services from government or charity organizations to address social and economic sustainability aspects of cancer prevention and treatment. The fifth macro-service is Information Availability, which emphasizes the need for the availability of information related to cancer prevention and treatment and includes three services, Breast Cancer Awareness, Awareness Campaigns, and Questionnaire and Competitions.

Subsequently, we show the possibility of finding additional services by topical searching over the dataset, where the topics could be macro-services, services, or other aspects of the services domain (cancer, in this case). We use four topical searches (Causes, Symptoms, Prevention, and Stakeholders) and detect 42 additional services that include Sports, Stress Avoidance and Control, and others. We call them extended services since these are discovered by using the knowledge gained from the initial service discovery process.

The methodology of discovering these services involves exploring the healthcare space related to cancer in Saudi Arabia using LDA-based topic modelling of Twitter data. The data shows the problems, solutions, strategies, activities, comments, and requirements of various stakeholders clustered into 20 themes that are used to develop 17 services and five macro-services by merging some clusters. The idea of developing cancer-related services from social media analysis is motivated by the well-known science that social media facilitates value co-creation through stakeholder interactions, allowing value creation or the creation of new services (see e.g., [16]).

We have developed a software tool from scratch for this work. The tool implements a complete machine learning pipeline including data collection, pre-processing, clustering, validation, and visualization components. The dataset we used contains over a million tweets (1,352,814 tweets to be precise) collected during the period 22 September–1 November 2021. We have tried to make the translation of the keywords and other Arabic content (sample tweets, etc.) contextual to allow English readers to understand the contextual use of the keywords. However, in other cases, we have used literal translation. Note also that we did not always quote a complete tweet for reasons of privacy, and in other cases as a part of the tweet was not relevant to the discussion. This work builds on our extensive research in social media analysis in English and other languages on topics in different sectors, see e.g., [17].

Novelty, Contribution, and Utilization

The literature review presented in Section 2 shows that while there are many works on the use of social media in healthcare, our work is novel in several respects. Firstly, none of the existing works on social media analytics in any language have focused on cancer as extensively as we do in this paper, discovering over fifty topics in several dimensions including cancer causes, symptoms, prevention, treatment, socioeconomic sustainability, and stakeholders. Therefore, this paper provides insight and evidence from social media public and stakeholder conversations about various aspects of the cancer disease in Saudi Arabia such as public concerns, patient requirements, solutions for problems related directly and broadly to cancer such as cancer treatment, financial difficulties, psychological ordeals and traumas, operational challenges for the families of cancer patients and other information. The fact that there are limited works in social media analytics in healthcare, particularly cancer, and considering that studies on social media analytics in the Arabic language within Saudi Arabia are even more limited, and that we have used a dataset that we carefully curated for this study differentiates our work from others.

Secondly, none of the existing works have proposed a similar approach of using social media and AI to extract healthcare or cancer services. We believe systemizing this approach could lead to a revolution in sustainable healthcare, driven by communities co-creating social values with the exchange of services for services that are not necessarily materialistic, but driven by equity, altruism, community strengthening, innovation, and social cohesion. Open service and value healthcare systems based on freely available information can revolutionize healthcare in manners similar to the open-source revolution by using information made available by the public, the government, third and fourth sectors, or others, allowing new forms of preventions, cures, treatments, and support structures.

The rest of this paper is organized as follows. Section 2 discusses the related work. Section 3 details the methodology and design of the Musawah tool. Section 4 introduces and explains the 17 discovered services and five macro-services. Section 5 discusses the 42 extended services that we discover by using knowledge gained from and extending the initial discovery process. Section 6 provides discussion. Section 7 concludes and gives directions for future work.

2. Literature Review

In this section, we provide a review on the topics related to this research paper. Section 2.1 reviews the works related to healthcare that have utilized social media. In Section 2.2, we review the literature related to the use of social media that focuses on cancer-related studies; Section 2.3 presents the same however limiting to works in the Arabic language alone. Section 2.4 reviews some works on automatic naming of topics and clusters. Section 2.5 discusses the research gap.

2.1. Social Media and Healthcare

Researchers have extensively used social media analytics in different application domains such as analysis of perception people have about AI and other technologies for city planning [18], city logistics [19,20], disaster detection [21], sentiment analyses of public opinions of government services [22], detection of various events related to road traffic [17,23,24], detection of symptoms and diseases nationally and across different urban areas [4], and urban governance during the COVID-19 pandemic [12].

Specifically, in healthcare, many researchers have used social media as a powerful tool for addressing healthcare system challenges and evaluating the conversations about various health-related topics. The COVID-19 pandemic has produced many challenges for public healthcare systems around the world. The main challenge is providing effective health care to patients while maintaining the safety of service providers on the frontline. Telemedicine appeared as a prominent service to achieve this end. This service integrates the use of technology into medical practices to reduce face-to-face meetings. Leite and

Hodgkinson [25] built a framework to examine how patients and doctors can co-create service value for the telemedicine ecosystem with/for any kind of crisis. They collected a sample about telemedicine and included 146 tweets from patients and 734 tweets from doctors. They applied inclusion and exclusion criteria on tweets to get the final sample. Thematic analysis was implemented to reveal seven themes in the data including equitable healthcare delivery and others. The result showed that the value co-creation framework is a platform for inclusive, sustainable, and equitable telemedicine services.

Some studies have used machine learning and deep learning methods for healthcare-related analytics using digital media data in healthcare. Jahanbin et al. [26] proposed an approach for extracting storing, monitoring, and visualizing data related to infectious diseases by using news and tweets mining techniques. The approach is called "Fuzzy Algorithm for Extraction, Monitoring, and Classification of Infectious Diseases (FAEMC-ID)". The proposed system collected 10,000 news items and tweets. Classification with the evolving fuzzy model is performed on the dataset. The classified data were visualized on the world map to identify the high-risk areas. The proposed system helped to monitor infectious diseases and control the spread of epidemics in a timely manner. The results of the proposed analysis showed a high accuracy of 88.41%, compared to other algorithms. Alotaibi et al. [4] constructed the Sehaa system, which is a big data analytics tool for improving healthcare in KSA using Arabic Twitter data. Sehaa tool consisted of four main modules. The data collection module captured 18.9 million tweets about health symptoms and diseases in KSA. The pre-processing module is used to clean tweets and manually label them by using two levels of labeling: 'related' and 'unrelated' tweets, and then labeling related tweets to 'awareness' or 'inflicted by' a disease. The third module was classification. This phase used six classifiers, including Naive Bayes, Logistic Regression, and four methods of feature extraction to detect different diseases in Saudi Arabia. The validation module was used to evaluate the performance of the classifiers by using F1-Score and accuracy. The results of the study were visualized and validated using external sources, including national statistics, news media, and research reports. The results revealed that (1) the top five diseases in KSA are dermal diseases, hypertension, diabetes, heart diseases, and cancer. (2) Riyadh and Jeddah needed more awareness about the diseases, while Taif was the healthiest city.

2.2. Social Media and Cancer

We review here the studies related to cancer where social media is used as the data source. First, we review works without regard to any language or a modelling method, and subsequently, in Section 2.2.1, we will discuss studies on cancer that have used topic modelling. Table 1 summarizes some of the studies that we have discussed in this section, including some other relevant works that, despite not having been discussed, we have summarized for the readers' interest.

The literature review has looked at various issues related to cancer disease. For example, some works have focused on studying emotions among cancer patients. Wang and Wei [27] constructed an approach to study the emotions shared by patients' cancer communities on Twitter. The study collected 53,026 tweets, posted by 39,504 Twitter users, which included 29,979 retweets and 23,047 original tweets. The deep learning (Recurrent Neural Network RNN) models were used to extract emotions from English tweets and classified them into anger, joy, sadness, fear, hope, and bittersweet. The results of the study showed that joy was the most used emotion in tweets, then sadness and fear, whereas anger, hope, and bittersweet were less shared. Also, influencers were among the account that posted the most content with positive emotions. Crannell et al. [28] built an approach that aims to analyze the content of tweets written by cancer patients in the US and compute the average happiness for patients based on the cancer type. The study collected 186,406 tweets from March 2014 to December 2014. The pre-processing step was implemented to clean the dataset. The tweets were filtered based on cancer diagnosis and using regular expression software pattern matching. They used different filters such as "Breast cancer", "Lung cancer", and others. The authors utilized the patients' Twitter identification numbers to collect all tweets for each patient and calculate the average happiness value by the quantitative hedonometric analysis and the Mechanical Turk (LabMT). The results of the study showed that the most common cancers were 'breast', 'lung', 'prostate', and 'colorectal'. The calculated happiness values for every type of cancer showed that the happiness values were higher for breast, thyroid, and lymphoma cancers, and lower for kidney, pancreatic, and lung cancers. The study proved that the patients express their illness on social media.

Several studies have focused on studying cancer types. A number of works have looked into breast cancer. Modave et al. [29] constructed an approach to understand the perceptions and attitudes of people related to breast cancer on Twitter. The study collected 1,672,178 tweets related to breast cancer. The pre-processing was implemented to clean and prepare tweets for classification models. The text was classified into three groups including irrelevant, promotional, and laypeople's discussions by using CNN and LSTM models in Keras library that run on top of the Tensorflow framework. Then, the sentiment analysis was performed by using the Linguistic Inquiry and Word Count (LIWC) tool to analyze emotions/attitudes into five statuses: sadness, anger, anxiety, positive emotion, and negative emotion. Finally, topic modeling used the Biterm algorithm to identify the main topics from relevant tweets. The detected topics include awareness month events, treatment, risk, family, diagnosis, news, friend, screening, pink goods, and study. The NLP with machine learning is a useful tool to evaluate people's attitudes from their health tweets and detect their perceptions about specific health topics. Meena et al. [30] built a system based on social media and Twitter to extract the sentiments for breast cancer disease and chemotherapy treatment. The study collected 10,000 tweets by using the keywords 'breast cancer' and 'chemotherapy'. The sentiment analysis was implemented using the Naive Bayes algorithm to classify the sentiment of tweets into neutral, strongly positive, weakly positive, strongly negative, negative, or weakly negative. Bigram analysis was used to find the most frequent word in the tweets for 'breast cancer' and 'chemotherapy'. The results of the proposed system showed that the sentiments for breast cancer and chemotherapy were weakly positive or neutral. Thus, sentiment analysis for tweets has a great impact on health care.

Rasool et al. [31] have developed a framework that aimed to predict breast cancer risk in real-time using ML and spark as big data analysis platforms. The study used 'Wisconsin Breast Cancer Data-Set (WBCD)', which was imported from UCI repository and included 699 records with 11 attributes. The preprocessing steps were used to remove unnecessary attributes and complete the absent values of the attribute. Five classifiers were used, to detect cancer to be benign or malignant, including Logistic Regression, Gradient Boosted Trees, Support Vector Machine, Random Forest, and Decision Tree. The performance of the five classifiers was evaluated by computing the accuracy, specificity, and sensitivity for algorithms. The results reveal that the highest performance obtained for predicting the risk of disease in real-time was by RF classifier. Diddi and Lundy [32] constructed a methodology that aimed to examine how the health organizations including "Susan G. Komen", "U.S. News Health", "Breast Cancer Social Media", and "Woman's Hospital", used their Twitter accounts to talk and support aspects of breast cancer in October, which is dedicated to breast cancer awareness. The study collected 2961 tweets, 136 of the tweets belong to Woman's Hospital, 180 tweets for US Health News, 280 tweets for Komen, and 2365 tweets for BCSM. After using qualitative coding, content analysis was used to analyze, and categorize the tweets. The result of the proposed system showed that the highest number of tweets was written by BCSM, then Susan G Komen, US Health News, and Woman's Hospital. Qualitative content analysis showed that the most prevalent construct of the Health Belief Model (HBM) was "perceived barriers", then "cues to action", "perceived benefits", "self-efficacy", and "perceived threat".

Some works have explored skin cancer. Silva et al. [33] aimed to analyze Twitter posts in Australia about skin cancer. The study collected 12,927 tweets. Data filtering was implemented by using inclusion and exclusion criteria, removing non-English, unrelated, or unclear content. The quantitative and qualitative analyses were implemented on datasets to calculate the number of tweets, finding the relationship between the number of tweets and temperature by using linear regression analysis. The content analysis identified three themes: health-related expressiveness, health information, and health advice. Nguyen et al. [34] developed a methodology that aimed to evaluate the influence of the campaign (Don't Fry Day) by the National Council on Skin Cancer Prevention (NCSCP)'s 2018 on Twitter, categorize types of accounts participating in the campaign, and determine the themes of the tweets. The dataset was collected from 1881 Twitter accounts. The study categorized the types of contributors in the campaign, based on username, into 'government-affiliated account (federal)', 'nongovernmental organization (NCSCP and health)', 'government-affiliated account (state/local)', 'health/cancer/medical center', 'individual', 'news/media', 'businesses', and 'other/unknown'. After manual coding of tweets, content analysis was used to identify themes of the tweets which are 'informative', 'minimally informative', and self-interest campaign promotion themes. The results of the study showed the large effective use of social media to promote public health campaigns and understand the types of messages and accounts involved in social media campaigns.

Table 1. Studies on the Types of Cancer Using Social Media.

Author	Type of Cancer	Tweets (×1000)	Description
Modave et al. [29]	Breast	1672	Understand the perceptions and attitudes of people related to breast cancer on Twitter.
Meena et al. [30]	Breast	10	Extracted the sentiments for breast cancer disease and chemotherapy treatment.
Diddi et al. [32]	Breast	3	Examined how the four health organizations used their Twitter accounts to talk and support aspects of breast cancer in October.
Silva et al. [33]	Skin	13	Analyzed Twitter posts in Australia about skin cancer.
Sedrak et al. [35]	Lung	26	Analyzed content related to lung cancer and examined the dialogues of lung cancer clinical trials.
Sutton et al. [36]	Lung	24	Examined the user type and the content of tweets of lung cancer to identify the nature of messages within the cancer control continuum.
Sedrak et al. [37]	Kidney	2	Implement an exploratory analysis for the content of Twitter related to kidney cancer and the participants in this content.
Nejad et al. [38]	Childhood	286	Examine how Twitter was used during childhood cancer awareness month (CCAM).

Some works have investigated lung cancer. Sedrak et al. [35] built a Twitter-based approach to analyze content related to lung cancer and examined the dialogues of lung cancer clinical trials. The study used Nvivo (qualitative data analysis software) to collect 26,059 tweets and prepare them for the analysis stage by removing duplicates and non-English tweets. The content analysis was used on the final sample to categorize tweeters as individuals, media, or organizations. The tweets were categorized into support, clinical trials, prevention, treatment, general information, diagnosis, screening, or symptoms. Most of the tweets were posted by individuals and were focused on support and prevention. The results of the study demonstrated that social media is a promising and useful source to explore how patients conceptualize and communicate about any health issues. Sutton et al. [36] constructed a methodology that aimed to examine the user type and the content of tweets related to lung cancer to identify the nature of messages within the cancer control continuum. The study collected 1.3 million tweets, including 23,926 messages related to lung cancer. The 3000 tweets were manually coded by three coders. Descriptive and inferential statistics were performed to study the content of tweets and types of users. The content of tweets was categorized into treatment, awareness, prevention and risk information, survivorship, end of life, diagnosis, active cancer–unknown phase, and early detection. The types of users were categorized into individual, organizational, media, and unknown. The SPSS was used to analyze data and chi-square was used to assess and identify the relationships between user types, and cancer content. The experimental result showed that most of the tweets focused on treatment, awareness, risk, and prevention topics. The majority of messages were tweeted by individual users. Twitter is a useful source in raising awareness about cancer by timely dissemination of information about the treatment and prevention methods.

Among the studies that have explored kidney cancer includes another study by Sedrak et al. [37], which have built an approach to implement an exploratory analysis for the content of Twitter related to kidney cancer and the participants in this content. The study collected 2568 tweets. The NCapture tool was used to collect tweets and save them into an Excel file. Non-English tweets and those not related to kidney cancer were excluded from the dataset. Finally, the content analysis methods were used to analyze the final dataset of 2097 tweets. The content of tweets was classified into different domains including support, general information, treatment, clinical trials, donation, diagnosis, and prevention. Also, the user was classified into individual, media, or organization. The results showed that 858 tweets were written by individuals, 865 tweets authored by organizations and 364 tweets written by media sites. Most discussed content included support by 29.3% and treatment by 26.5%.

Some studies have looked into the types of children's cancer. Nejad et al. [38] built a descriptive-analytical methodology to examine how Twitter was used during childhood cancer awareness month (CCAM). The study collected 285,859 tweets related to childhood cancer, which coincided with CCAM. Latent Dirichlet allocation (LDA) was used as a method of content analysis to discover the topics using MALLET from tweets. The detected topics included awareness, loved ones, walks and runs, wearing gold, and fundraising. Twitter user was categorized into four groups including news agencies, individuals, organizations, and celebrities. They observed that the most frequent topics are awareness and fundraising. The experimental results for the proposed system showed that Twitter has proven to be an effective channel for communication and raising awareness about childhood cancer. Thus, Twitter can play an important role in the dissemination of awareness among patients.

Some studies have focused on cancer treatment using social media data to utilize patients' experiences with cancer treatment methods, including chemotherapy and radiation. Meeking [39] built an approach to understanding the role of social media platforms like Twitter, as a novel digital data source, in understanding people's experiences of radiotherapy. The study collected 442 unique tweets about radiotherapy written by patients and their families. The quantitative content analysis was used to classify the data based on the content of the tweets or the user Identity into different types of accounts such as patient, healthcare organization, and healthcare professional. Thematic analysis was used to detect six themes. Three main themes related to the pathway of radiotherapy: pre-, during-, and post-treatment. The other three themes included 'emotional and informational support', 'impact on loved ones', and 'giving thanks'. The paper highlighted the psychological and physical effects of treatment and how patients seek to get emotional and information support from social media. The informational support can be improved by increasing online support. The experimental results show that Twitter is a useful platform to share and discuss people's experiences with radiotherapy. Other studies on utilizing social media applications such as Twitter for cancer treatment include [40].

2.2.1. Topic Modeling and Cancer

We review in this section the works about cancer analysis using social media that have utilized topic modelling algorithms such as Latent Dirichlet Allocation (LDA). Zhang et al. [40] aimed to examine and compare patients' and healthcare professionals' perceptions of chemotherapy by analyzing chemo-related tweets. The study gathered tweets by using the keywords "chemotherapy"or "chemo" from the cancer-related accounts. They collected 13,273 individual tweets and 14,051 organization tweets. Topic modeling, sentiment analysis, and a word co-occurrence network were used to analyze the content of tweets. LDA was used to detect the most important topics related to chemotherapy, which are hormone therapy, radiation therapy, and surgery. A word co-occurrence network was shown to have a strong relationship between chemo-surgery tweets. Sentiment analysis was used to determine how patients felt about chemotherapy outcomes and their attitudes toward chemotherapy across disease types. The study shows that Twitter is a helpful healthcare data source for oncologists (organizations) to better understand patients' experiences during chemotherapy, generate personalized treatment plans, and enhance clinical electronic medical records (EMRs). Nejad et al. [38] examined how Twitter was used during childhood cancer awareness month (CCAM). They applied LDA and extracted six topics from 285,859 tweets including awareness, clothing, loved ones, walks, fundraising, and art. Modave et al. [29] constructed an approach to understand the perceptions and attitudes of people related to breast cancer on Twitter. They applied Biterm topic modeling on 1,672,178 collected tweets related to breast cancer. The detected topics include awareness month events, treatment, risk, family, diagnosis, news, friend, screening, pink goods, and study.

2.3. Social Media and Cancer (Arabic Language)

There are some studies related to cancer analysis using Twitter in the English language, however, the research studies about cancer using social media in the Arabic language are limited. We have found only one such work on social media analytics in cancer research, which is relevant to this paper. The work is by Alghamdi et al. [41] and is focused on studying chemotherapy misconceptions among Twitter users. They built an approach to evaluate Twitter conversations and find misconceptions about chemotherapy among Arabic populations. They collected 402,157 tweets by using the Twitter Archiver tool. Retweets and irrelevant tweets have been removed. A manual content analysis was implemented to classify the users, themes, and common misconceptions for chemotherapy. The categories of the user included cancer patients, general users, relatives/friends of patients, accounts for health, accounts for Media, and cancer specialists. Tweets' themes categories included advice and information, misconception, experience, prayers and wishes, seeking or offering medical/financial help, seeking medical information/advice, and analogy. Statistical analysis was implemented by using the SPSS software and reported that the most of tweets were written by general users, then followed by the friends and relatives of cancer patients. For themes, prayers and wishes were the most popular topics. Descriptive statistical analysis was performed to find the frequencies between every theme and the user category. The chi-square test was used and the result revealed that a high association exists between the themes of the tweets and the category of the users.

2.4. Automatic Naming of Topics

The automatic naming of topics is desired and has been researched in the literature. We discuss a few works here to introduce this topic to the reader. The probabilistic topic models aim to reveal latent topics from the corpora of text. Probabilistic topic models represent each document as a set of topics and every topic has a set of words. Humans can label every topic based on the top words of the topic, however, difficulties arise when they do not have good knowledge in the field of the documents. To address the problem, automatic topic labeling techniques can be used to assign interpretable labels for topics. Allahyari et al. [42] built a knowledge-based topic model, namely, KB-LDA that combines topic models with ontological concepts. The study aimed to use the semantic knowledge graph of concepts in ontology and their diverse relationships, as DBpedia, with unsupervised topic models (LDA) to automatically generate meaningful topic labels. The proposed approach helped to describe topics in a more extensive way and improve the quality of topic labeling by making use of the topic-concept relations, which can automatically generate meaningful labels. Lau et al. [43] constructed an approach for automatically labeling topics produced from the LDA topic model. The study generated a label candidate set from top-ranking topic terms, Wikipedia titles containing the top topic terms, and sub phrases extracted from

Wikipedia articles. Then, it used the ranking method to find the best label for each topic by using a mixture of association measures and lexical features, fed into a supervised ranking model using the support vector regression (SVR) model.

Wan and Wang [44] proposed a method that aimed to use text summaries for automatic labeling of topics produced by topic models. The study extracted several sentences from the most related documents to form the summary for each topic. The system proposed a summarization algorithm based on submodular optimization to create summaries with high coverage, relevance, and discrimination for all of the topics. The results of the proposed method showed that the use of summaries as labels had apparent advantages compared with the use of words and phrases. He et al. [45] built a graph-based ranking model, namely TLRank, that aimed to automatically label every topic produced from topic models. The proposed model used the strategy of enhancing matrix transition probability based on the textual similarity between vertices and the characteristics of vertices (sentences). The strategy seeks to restrain the topic label redundancy and enhance its diversity. The experiment results showed that the TLRank model has the ability to generate topic labels with high coverage, relevance, and discrimination. Suadaa and Purwarianti [46] introduced a methodology to cluster and label the Indonesian text by using Latent Dirichlet Allocation (LDA) and Term Frequency-Inverse Cluster Frequency (TFxICF). The methodology performed text preprocessing processes by abbreviations extraction, tokenization, stemming, and stop-words removing. For topic modeling, Latent Dirichlet Allocation (LDA) is implemented to cluster documents and detect hidden topics. The clustering quality was evaluated by using precision, recall, and F-measure. For the labeling phase, the study has used Term Frequency-Inverse Cluster Frequency (TFxICF) to label each cluster based on the phrase or word tokens that have the highest TFxICF.

2.5. Research Gap

The literature review presented in this section establishes that while there are many works on the use of social media in healthcare, our work is novel as none of the existing works in any language have focused on cancer as extensively as we do in this paper and, secondly, none of the existing works have proposed a similar approach of using social media and AI to extract healthcare or cancer services.

3. Methodology and Design

This section explains the methodology and design of the Musawah tool. The proposed system architecture is described in Figure 1. It contains five components: data collection, pre-processing, service discovery, visualization, and evaluation; these components will be discussed in Section 3.3 to Section 3.7, respectively. Before we describe the system architecture, we discuss the conceptual development of the Musawah tool and system in Section 3.1 and give an overview of the tool in Section 3.2.

3.1. Musawah Service Co-Creation System: Developing the Concept

This paper proposes a data-driven artificial intelligence (AI) based approach called Musawah to automatically detect and identify healthcare services that can be developed or co-created by various stakeholders using social media analysis. We called our approach and system Musawah (equity in Arabic) as we believe that the current healthcare trends and doctrines are not sustainable—socially, environmentally, and economically—and that, with a belief and practice in equity by people, our approach of open service and value healthcare systems based on freely available information can enable equity and sustainability. The case study focuses on cancer disease in Saudi Arabia using Twitter data analytics in the Arabic language; however, the proposed approach can be used broadly for any disease or purpose and in any language. We discuss first the conceptual development of the Musawah tool and system in the following discussion where topics include the need for data sharing and participatory governance, developing services, value and service co-creation, and the motivations behind it.

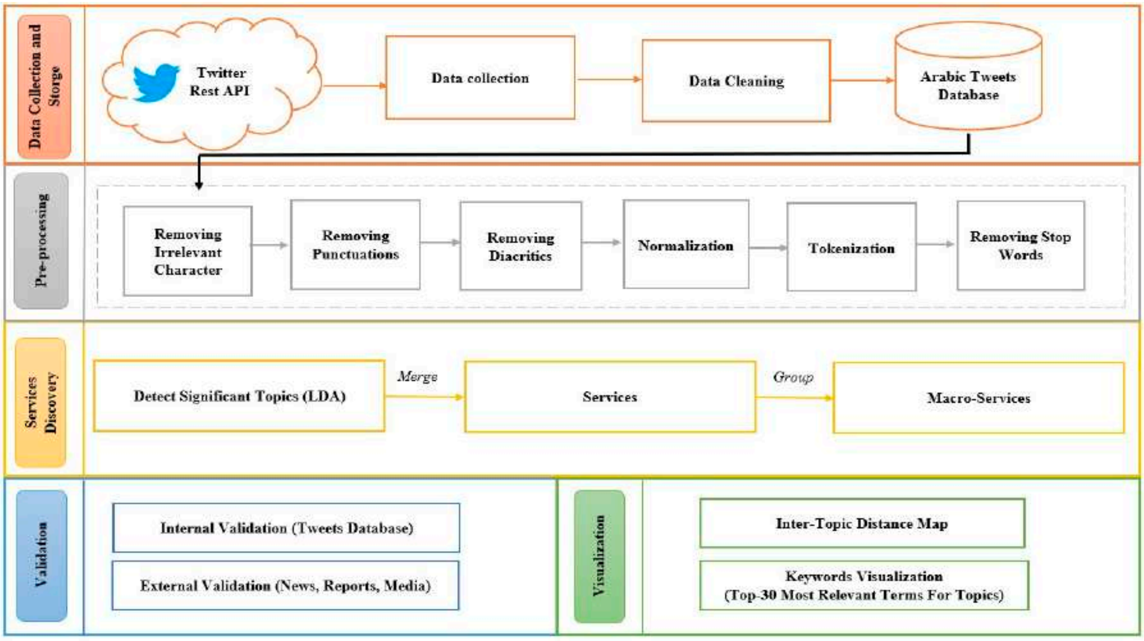


Figure 1. The Proposed System Architecture.

Making cities smarter needs a holistic approach to manage, improve, and adapt the services provided by the city to its citizens. A smart city can be seen as a system of services [47]. With this vision, services are the basic concept of smart cities. Service science studies value co-creation as an interaction mechanism between technology, people, shared information, and organizations to improve operational efficiency, enhance citizen welfare and the quality of government services. The focal point for this vision is based on the co-creation of value through the direct citizens' participation in the service development and evaluation process. In a smart city, value co-creation is achieved by data sharing and the exchange of knowledge between citizens and the city [48]. It is necessary to understand smart service systems to foster innovations and the development of these systems in different fields. This will help to enhance understanding among people, facilitate collaborative analysis, create synergy between applications, and support systems development for smart cities.

Governments can utilize open data as a new method for developing services that allow external stakeholders to increase their role in the innovation of government services [49]. This is in contrast to previous methods of e-government service innovation, the governments create and develop services by the agencies themselves. Currently, organizations can achieve open service innovation by including the external stakeholders in enhancing existing services and developing new services based on comments or ideas generated by their stakeholders and from the problems their encounter. The open innovation in services refers to the transition from the closed innovation model based on internal knowledge to the innovation paradigm that resides both externally and internally to an organization [49]. For example, the Service Systems Development Process (SSDP) [50,51] contains a set of iterative phases that require inputs and expected outputs for every phase to understand the activities and needs required for realizing service systems. These processes are implemented as an iterative approach with the alignment of service system entities to understand the impact on enterprise capabilities and processes, technology, information systems, and customer expectations. Firstly, in the service strategy stage, newly developed services are determined by examining and knowing end-user needs, organization strategies, trends of mass collaboration, and trends of technology. The organization decides to develop the selected service systems based on the high-level socio-techno-economic feasibility study. During the service design and development stage, the requirements are analyzed. The service system entities' functions and linkages are identified. Also, it ensures the achievement of the activities of service integration, verification and validation, which include information exchange between the entities of the service system to provide the service continuously. In the service transition and deployment stage, the service is tested to ensure readiness insertion and operation. When the service is deployed, it enters into the operations stage [50,51].

Governments are making increased efforts to innovate public service delivery, such as under the e-government umbrella [52]. The success of e-government depends on the need to better understand the requirements of citizens and include them in the development of e-government services. This can be achieved through using the service systems perspective as “a guiding framework” to analyze the development initiatives for participatory e-government services. Specifically, these initiatives can be analyzed as a service system based on the four key resources of people, shared information, organizations, and technologies, which will help to derive possible developments and enhancements for services [52]. Similar concepts have been discussed by many researchers in the past. For example, Salminen [53] has proposed the new service development process (NSD) that contains five main stages. The most important component of the model is the continuous interaction with the customer. This interaction is dependent on the common value model, which helps companies to discuss with the customer the potential value of the new services such as cost-effectiveness, process improvements, or new product features. There are three phases for the value management process; “Value Evaluation/ Assessment”, “Value Creation”, and “Value Maximization/Delivery” [53].

Health care affects economies dramatically worldwide, while affecting directly the quality of life of individuals. Traditionally, in healthcare systems, customers have been seen as passive recipients of services instead of being active elements. With rising healthcare expenditures and a growing desire for more personalized and better care, healthcare systems have recognized the importance of patients in co-creating the healthcare service experience during the last decade [54]. The Service-Dominant Logic (SDL) concept captures and represents this new paradigm of co-creating [55–57]. Note that SDL is a general concept and paradigm and is not limited to healthcare. SDL could be used to incorporate patients in creating value and producing services in this environment. Value co-creation is defined in this approach as a process, in which several actors exchange resources and collaboratively create and produce value. As a result, the information sharing between the actors of services is critical to SDL and value co-creation [16]. The new technologies represent an important role in facilitating the process of value co-creation in service science. We can build a more connected and smarter healthcare system that can provide greater help, predict and prevent illness, enable patients to make more responsible decisions. The emergence of social media, in particular, has resulted in significant changes in the processes of creating, promoting, and sharing content. The use of social media by people and public health organizations is being used to communicate in new ways for various mutually beneficial activities [54].

Note that while these and other works have investigated the use of social media for value, co-creation by various actors communicating to each other, and sharing information, none of the works in the literature have proposed the discovery of healthcare services automatically from social media. This work, therefore, contributes a novel approach of using social media and AI to discover, develop, and exchange healthcare or cancer services. We demonstrate the potential of our approach by discovering over fifty services in several dimensions of cancer disease including cancer causes, symptoms, prevention, treatment, socioeconomic sustainability, and stakeholders. We plan to systemize this approach by creating open service and value healthcare systems based on freely available information made available by the public, the government, third and fourth sectors, or others, allowing new forms of preventions, cures, treatments, and support structures to be discovered and exchanged between various parties—an exchange of values for values or services for services that are not necessarily materialistic but driven by equity, altruism, community strengthening, innovation, and social cohesion. Service co-creation is enabled through engaged participation of need and value propositions. In co-creation terminology, both parties, a customer and company, create value for each other and this value could be money, product, service, or anything of value for the parties involved. Therefore, the value could be anything materialistic or otherwise, such as a desire for equity, the altruistic nature of a person or party, and the desire to strengthen communities, bringing innovation, or social cohesion. Moreover, our approach also allows the gaining of insight and gathering of evidence from social media public and stakeholder conversations about various aspects of the cancer disease in Saudi Arabia such as public concerns, patient requirements, solutions for problems related directly and broadly to cancer such as cancer treatment, financial difficulties, psychological ordeals and traumas, operational challenges for the families of cancer patients and other information.

As regards the discovery and naming of services, in this paper, we used a process of discovering clusters or topics (we call them services) using topic modelling of Twitter data using the Latent Dirichlet Allocation (LDA) algorithm. We experimented with discovering a different number of clusters using LDA—10, 15, 20, and other numbers—and found that 20 topics gave the best results for discovering important services from Twitter data. The number of clusters to be detected is dependent on the size, nature, and other properties of data, and this is an active area of research. Following the extraction of topics, we merged some of these topics, while others were kept individually. We call these merged or individual topics 'services', and group these services further into groups of services called 'macro-services'. This whole process created a set of 17 services and five macro-services. The services were named manually using topic keywords and domain knowledge. We have provided a review of the automatic naming of topics (services in this case) in Section 2.5. This is an area that we plan to work on in the future to automate the service discovery, definition, transition, and deployment process.

More details on the methodology will follow in the rest of this section and the paper. The paper has contributed to certain areas, while other topics such as automatic naming of services, will be investigated in the future.

3.2. System Overview

We built the Musawah system in order to discover healthcare services that can be developed or co-created by various stakeholders using social media analysis. The proposed approach can be used for any disease or purpose and in any language. However, this work focuses on cancer disease in Saudi Arabia using Twitter data analytics in the Arabic language. The proposed system architecture contains five components: data collection; pre-processing; service discovery; visualization; and evaluation. First, we collected and downloaded tweets using Twitter REST API and the Tweepy library with the set of predefined parameters as Keywords and the language. The tweets collected were tweet objects in the JSON (JavaScript Object Notation) format. After collecting JSON files, we converted and saved them into CSV format. It is well known that the contents of social media contain noise and unstructured data, and many errors. Therefore, in the Pre-Processing component, the acquired data are cleaned and preprocessed to ensure its readiness for the analysis stage. For removing stop-words, we used the Natural Language Toolkit (NLTK) library with add a new list of stop-words to remove from the text. Also, we used the simple_preprocess() in Python to tokenize each tweet into a list of words and used a regular expression in python to remove the unnecessary characters from the tweets. Subsequently, we used the Gensim Python library to build a service discovery module using Latent Dirichlet Allocation (LDA). We have adjusted the settings of the parameters of the LDA model: the number of topics was 20, passes were 10, and iteration was 100. The number of topics is considered an important parameter for building a model. The passes refer to how many times the algorithm is supposed to pass over the whole corpus. The iterations are the maximum number to iterate every document in the corpus for calculating the probability of each topic. Then, we visualized the topics by using the LDAVis tool that provides a web-based interactive visualization of topics generated from LDA. Finally, the results are validated by the detected services using two techniques, which included external and internal validation.

3.3. The Dataset

The data was collected using Twitter REST API. Tweets were collected using various terms of cancer, its types, as well as its examination and treatment methods. For example, we used the keyterms “سرطان” (Cancer), “ورم” (Tumor), “ورم خبيث” (Malignant), “سرطان لرئة” (Lung Cancer), “الخزعة” (Biopsy), and others. Furthermore, we used several hashtags related to the types of cancer, its treatment, and awareness such as "سرطان الرئة#" (# Lung Cancer), "سرطان الثدي#" (#Breast Cancer), “سرطان البروستات#” (#Prostate Cancer), and others. In Table 2 we provide a sample of the keyterms and hashtags that were used for collecting data. A total of 1,352,814 tweets were retrieved during the period from 22 September to 1 November 2021. The tweets collected were tweet objects in the JSON (JavaScript Object Notation) format. After collecting JSON file, we converted and saved them into CSV format. Then, we cleaned the dataset by removing duplicates.

Table 2. The Keywords and Hashtags used to collect Dataset.

Keywords
السرطان، ورم، خبيث، سرطان الرئة، خزعة، طفرة جينية، الكشف المبكر، لوكيميا، الكيماوي، المسرطنة
Cancer, Tumor, Malignant Lung Cancer, Biopsy, Genetic Mutation, Early Screening, Leukemia, Chemo, Carcinogenic
Hashtags
#سرطان_الرئة، #سرطان_الثدي، #سرطان_البروستات، #افحصي_الان، #مرضى_السرطان، #علاج_السرطان، #العلاج_الكيماوي، #الكشف_المبكر، #مكافحة_سرطان_الثدي، # مرض_السرطان
Lung Cancer, Breast Cancer, Prostate Cancer, Check-up Now, Cancer Patients, Cancer Treatment, Chemotherapy, Early Detection, Fighting Breast Cancer, Cancer Disease

3.4. Data Pre-Processing

Data preparation or preprocessing is a necessary step to complete the steps of the data analytics process. Social media generates unstructured and informal data. This involves implementing different techniques to clean the acquired data. To ensure data readiness, data preprocessing should be performed to prepare the acquired data for subsequent steps of the analysis. This will increase the accuracy and quality of data analytics. The main steps of data preprocessing include: removing irrelevant characters; tokenization; normalization; and removing stop-words.

3.4.1. Removing Irrelevant Characters

The collected tweets may contain duplicates. To prevent having the same tweets, we removed the duplicates out of the collected tweets after saving and loading tweets in DataFrame format. Cleaning data included removing emails, redundant spaces and lines, single quotes, repeating characters, English alphabets, and all traces of emoji from a text file. Moreover, we removed the punctuation marks remove by creating a list for all the marks of punctuation, such as: [:, ., ?, ;, ؟, %, @, &,], and other types of brackets, mathematical, and slashes symbols. This will decrease the size of the feature set and make information more valuable. We also removed eight diacritics marks:

- Three diacritical marks that are used to refer to the short vowels: Fatha (اَ), Kasra (اِ), Damma (اُ);
- Three double diacritic marks: Tanwin Fath (اً) , Tanwin Kasr (اٍ) , Tanwin Damm (اٌ) ;
- Single diacritical mark to refer the absence of a vowel: Sukun (اْ) ;
- One diacritical mark to refer the duplicate occurrence of a consonant: Tashdid (اّ) .

3.4.2. Tokenization and Normalization

We used the simple_preprocess() method in Python to tokenize each tweet into a list of words. Normalization transforms words to a basic and consistent form. Normalization of Alef, Yaa, Hamza, and Taa Marboutah, in which Alef (أ آ إ ا) was replaced with (ا), Yaa (ي ى) was replaced with (ي), and Taa Marboutah (ة ه) with replaced with (ة).

3.4.3. Stop-Words Removal

Stop-words are distributed between the texts, are not very useful, and can be effectively removed. The role of removing stop-words is to delete the word that is not important in extracting the information. The tweets can contain stop-words like: "الى", "عن", "من". We used the Natural Language Toolkit (NLTK) library with add a new list of stop-words to remove from text, such as ("منكم", "لدي", "اي", "كان", "انها", "او", "والله", "ان", "ثم", "الى", "بينما", "بيني", "بشيء", "ممكن", "اذا", "وين", "اني", "لان", "حتى", "شي", "قبل", "يتم").

3.5. Topic Modeling Using Unsupervised Machine Learning

Academic studies and research rely on the use of computerized analytics to understand large volumes of unstructured text that cannot be analyzed manually due to the limitations of human data processing. In this area, topic modeling is a common technique used for data analysis and topics discovery. The topic model can be defined as "a collection of algorithmic approaches that seek to find structural patterns within a collection of text documents, producing groupings of words that represent the core themes present across a corpus" [58]. In this paper, we used Latent Dirichlet Allocation (LDA) as one of the types of topic modeling. LDA is a statistical model used to determine the important topics discussed in a set of documents (tweets in our study). It creates models in an unsupervised mode, which means no need for label training data. Every document is characterized by different probability distribution among different topics. The topics are formed based on the co-occurrence of keywords with a certain probability in the same document. The LDA model takes the collection of documents as input and generates a set of topics as the output. Every topic has a set of keywords in different proportions as well. The algorithms of topic modeling are not able to name the topic. So, the topics can be labeled by humans using the keywords of the topics.

We used Genism for LDA analysis in the Python environment. The tweets represent the documents in this work. We have adjusted the set of the parameters of LDA model: the number of topics was 20, passes were 10, and iteration was 100. The number of topics is considered an important parameter for building a model, the large number of topics requires a lot of overfittings, while the small number of topics makes the model underfitting. The passes refer to how many times the algorithm is supposed to pass over the whole corpus. The iterations are the maximum number needed to iterate every document in the corpus for calculating the probability of each topic.

3.6. Visualizations

In this paper, we used LDAVis to visualize and label the topics. LDAVis is a web-based interactive visualization of topics generated from LDA. It provides an overview of the topics by displaying them in the circles' form. It also displays the words closely related to each topic, and the degree of relevance of each word to topics. The visualization contains two basic pieces. In the left panel, the topics are visualized as circles. In the right panel, it has a horizontal bar chart that represents the most useful individual terms for interpreting the currently selected topic on the left.

3.7. Evaluation and Validation

The proposed system followed two techniques to validate the detected services: external and internal validation. We used different online sources and news media as an external technique for validation. The internal technique for validation is based on the collected Twitter data, which gives detailed information such as the account that was used to post the information, the time, and date of posting the tweet.

4. Results and Analysis

This section introduces and explains the discovered services and macro-services. Section 4.1 provides an overview of the macro-services and services detected from the Twitter data using the LDA algorithm. This is followed by five subsections, Section 4.2 to Section 4.6, with an explanation for each service with examples from Twitter posts.

4.1. Services and Macro-Services: An Overview

We discuss in this section the results of the system we propose. We describe the approach of discovering 17 services from Twitter data using the Latent Dirichlet Allocation algorithm (LDA). The approach involved extracting 20 topics by utilizing the LDA algorithm on Twitter data, and then combining a few of these topics into topic clusters which we call services. Throughout this research paper, the terms such as topic cluster, cluster, and service will be used interchangeably. In addition, we have grouped these seventeen services into five macro-services. In the previous section, we described the method for detecting these topics using the LDA algorithm. Python is used to develop the software.

Table 3 provides a list of services along with their corresponding data. Column 1 of the table lists the five macro-services. They are Prevention, Treatment, Psychological Support, Socioeconomic Sustainability, and Information Availability. In Column 2, we list seventeen services such as Early Diagnosis, Prevention and Control, Causes, and so on. Every macro-service contains one or more services. For instance, the second macro-service (Treatment) contains two services, which are Chemo and Radiation Therapy, and Surgical Therapy. In Column 3, the services numbers are listed. We have already discussed that we extracted 20 topics from Twitter data using LDA, and some of those topics that are related to one another are merged to form services. As an example of the merging of topics is the service (Breast Cancer Awareness), which is the result of merging Topic 9 and Topic 20. Column 4 shows the keyword percentage of the services in the table. Among the keywords in total, Topic 1 contains 9.7% out of the total keywords. Topic 2 contains 7% of the total keywords. Column 5 presents 10 keyterms (in total for all cluster topics) for each of the services. We initially collected 30 keywords per topic. From the initial 30 keywords, 10 keywords were manually chosen (using domain expertise) based on their importance to the relevant topics. Each keyword is listed in Arabic, along with its translation into English. These keywords and other content in Arabic language (example tweets, etc.) were translated contextually so that English readers can better understand the context of the keywords.

We extracted a taxonomy for the 17 services that were detected by our tool (See Figure 2). The taxonomy was created from Table 3, and it shows the services, their macro-services, and some keyterms associated with the services. The first level represents the macro services e.g., Prevention, Information Availability, and Treatment. Every macro-service contains one or more services. The second level branches represent these services e.g., Symptoms, Early Diagnosis, Causes, and Screening. Each service is characterized by various keywords. These keywords are provided in the third level branches. For instance, the keyterms Malignant, Examination, and Transform represent the Symptoms service.

Table 3. Macro-Services and Services Discovered from Twitter Data.

Macro-Services	Services	Topics	Keywords (%)	Keywords
Prevention	Early Diagnosis	1	9.7%	المبكر، الكشف، الفحص، الإصابة، اعراض، الوقاية، للكشف، نسبه، الشفاء، بادري Early, Detection, Examination, Injury, Symptoms, Prevention, Detection, Ratio/Percent, Healing, Go Early
	Prevention and Control	2	7%	سرطان، خبيث، الحياة، خير، للوقايه، اشياء، ورم، تأخذ، اسالكم، مرض Cancer, Malignant, Life, Good, Preventative, Things, Tumor, Take, Request, Disease
	Causes	15	4%	اثناء، بسبب، مرض، معرض، تابعه During, Because, Disease, Exposed, Related
		16	3.6%	مزمن، جينيه، طفره، ورم، الأشعة Chronic, Genetic, Mutation, Tumor, Radiation
	Screening	17	3.3%	الفحص، الذاتي، الماموجرام، بجهاز، دقايق، ينقذ، ورم، صحة، عيادة، جيت Examination, Self, Mammogram, Device, Minutes, Save, Tumor, Health, Clinic, Came
	Symptoms	19	3%	سرطان، بمراكز، بالفحص، وتتحول، احد، يجعلها، خبيث، مريضه، للثدي، اسالنكم Cancer, Centers, Examination, Transform, One, Make, Malignant, Sick, For the Breast, Appeal You
Treatment	Chemo and Radiation Therapy	3	6.6%	يحتاج، الاورام، علاج، المجاني، الكيماوي، الم، لمدد، تلقي، الحالة، استخدام Need, Tumors, Treatment, Free, Chemotherapy, Pain, Duration, Receive, Case, Use
	Surgical Therapy	4	6.3%	خبيث، استئصال، المقطع، ورم، يجب، عمليه، نوع، جانب، سرطاني، وجراحه Malignant, Resection, Part, Tumor, Must, Operation, Type, Side, Cancerous, Surgery
Psychological Support	Spiritual Support	5	6.2%	اللهم، مرضي، يارب، مرض، اشف، الم، اجسادهم، مرضى المسلمين، سقما، الشافي Oh God, Patients, Lord, Disease, Heal, Pain, Their Bodies, Muslim Patients, Sick, Healer
	Suffering	14	4.1%	مرض، رحمه، الخير، انسانيه، دعواتكم، اثر، حاله، معاناه، القاتل، صراع Disease, Mercy, Goodness, Humanity, Your Prayers, Impact, Condition/Case, Suffering, Killer, Struggle
	Hope and Optimism	18	3.1%	يدعي، اللهم، ياتي، بجميع، امنياته، محمل، اليكم، يعود، يا رب، جعلت Pray, God, Come, With All, Wishes, Loaded, To You, Return (Health), Lord, Made

Table 3. *Cont.*

Macro-Services	Services	Topics	Keywords (%)	Keywords
Socioeconomic Sustainability	Government Support	6	5.7%	الصحة، عاجله، بأهمية، تقيم، خدمه، المنطقة، وزاره، ادويه، مرضى، الألات Health, Urgent, Important, Evaluation, Service, Region, Ministry, Medicine, Patients, Machines
	Socioeconomic and Operational Challenges	8	5.4%	مريض، العلاج، بحاجه، اسال، جراحه، الف، فاتوره، حاله، تعاني، رقم، الرجاء Patient, Treatment, Needs, Ask, Surgery, Thousands, Bill, Case, Suffering, Number, Please
	Charity Organizations	11	4.6%	سنه، سند، توقع، عاجل، مريضه، نداء، زهره، حاله، قررت، انساني Year, Sanad Association, To Sign, Urgent, Patient, Call, Zahra Association, Case, Decided, Humanitarian
	Financial Support	12	4.5%	سرطان، ولدي، علاج، الفاتورة، حاولنا، معسر، ما نقدر، العملية، اناشدكم، خزعه Cancer, I Have, Treatment, Bill, We Tried, Poor, We Can't, Operation, Exposure, Begging, Biopsy
Information Availability	Breast Cancer Awareness	9	5.3%	الصحية، بالتعاون، حمله، جمعيه، مركز، للتوعية، محاضره، افحصي Healthy, Cooperation, Campaign, Association, Center, Awareness, Lecture, Examination
		20	2.9%	الوردي، المراه، باللون، بنشر Pink, Woman, Color, Spreading
	Awareness Campaigns	10	4.8%	حساب، السوال، اكتوبر الوردي، رتويت، الرد، بهشتاق، التغريده، الصحي، شعار، الحمله Account, Question, Pink October, Retweet, Reply, Hashtag, Tweet, Healthy, Logo, Campaign
	Questionnaires and Competitions	13	4.4%	متابعه، المشاركه، الأمان، انتشرت، شارك، فعالية، عضو، الإجابة، عام، الاف Follow, Participation, Safety, Spread, Share, Event, Member, Answer, General, Thousands

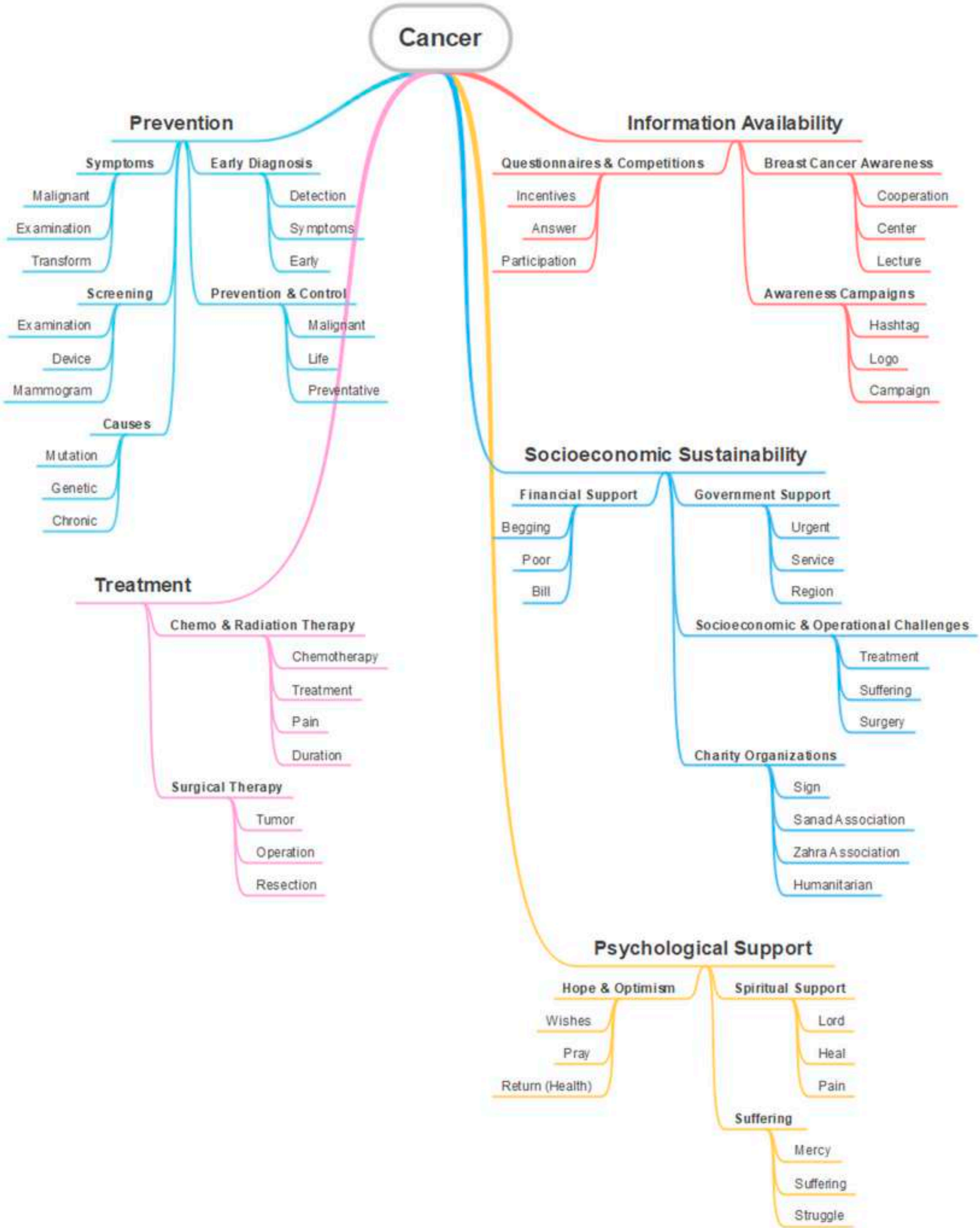


Figure 2. A Taxonomy of Discovered Cancer Services.

After explaining Table 3, we now move on to the explanation of topics using graphical information. Figure 3 shows the inter-topic distances among the extracted 20 topics based on the multidimensional scale. This figure shows the topic sizes in the bottom-left corner. Topic 1 is depicted by the largest circle, reflecting that Topic 1 is the largest topic based on its number of keywords (9.7% see Table 3). Figure 4 shows the top 30 most relevant terms (or keywords) for Topic 1. These terms are in Arabic language. Table 3 provides the English translation of the keyterms. The keywords are arranged in a decreasing order of their frequency within Topic 1 (represented by maroon bars). Each keyword has a blue bar, which represents the overall term frequency. After discussing the table and topic diagrams in general terms, now we will proceed to discuss in detail each of the services with data collected from the tweets and external sources. The discussion will also present additional detail on the topic diagrams for service.

Figure 3. The Intertopic Distance Map of the Clusters.

4.2. Prevention

We begin discussing the services related to the first macro-service, Prevention. It involves actions that reduce the chance of getting cancer (e.g., maintaining a healthy lifestyle, taking vaccines, and avoiding cancer-causing substances) and it also involves strategies and procedures that prevent cancer from getting worse (e.g., early detection, and regular screening tests). This macro-service includes five services, Early Diagnosis, Prevention and Control, Causes, Screening, and Symptoms. The first service is Early Diagnosis (see Row 1, Table 3) and is represented by keyterms from Topic 1 such as Early, Detection, Examination, Symptoms, Prevention, Treatment, and Healing. Figure 4 depicts the top 30 most relevant terms for Topic 1 (since there are 20 topics to cover, we are unable to include these figures for all of them, to avoid an excessive number of figures and to comply with the publisher's guidelines). Table 4 gives the English translation of the Arabic keywords in the figure. Early diagnosis of cancer is important for preventing bigger problems associated with health (e.g., cancer progressing to advanced stages), treatment options, and treatment costs. It focuses on detecting symptoms as soon as possible. Therefore, it can contribute to the detection of the cancer disease in its early stages. Early detection of cancer is very important for preventing reaching late stages, facilitating treatment, and accordingly, reducing the risk of death. For this purpose, it is one of the strategies used for successful treatment, and also for lowering treatment costs. The tweets related to this topic were about the early diagnosis of cancer and its importance. They were posted by official accounts, news accounts, and medical accounts. For example, the Saudi Cancer Society posted the following tweet about the importance of early detection of cancer in increasing the cure rate.

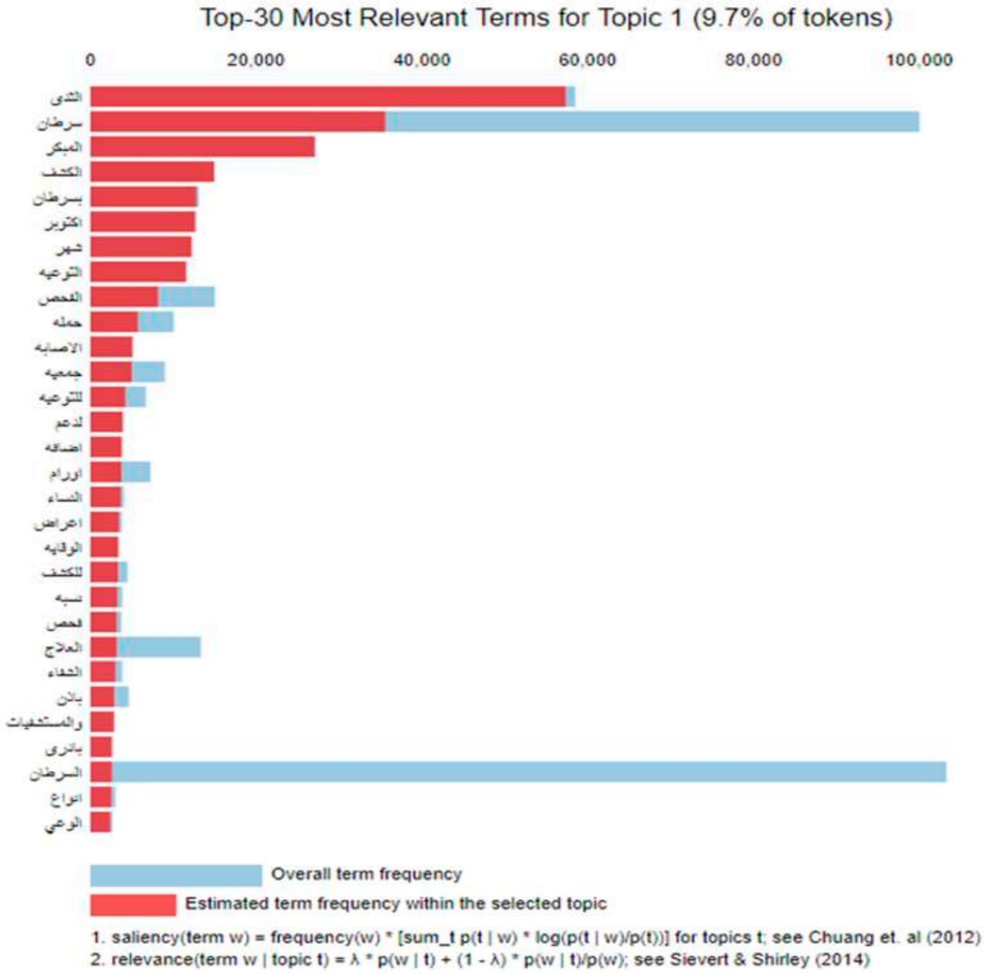


Figure 4. The Top 30 Most Relevant Terms for Topic 1.

Table 4. The keywords and their English translation for Figure 4.

Keywords	الثدي، سرطان، المبكر، بسرطان، اكتوبر، شهر، التوعيه، الفحص، حمله، الاصابه، جمعيه، للتوعيه، لدعم، اضافه، اورام، النساء، اعراض، الوقايه، للكشف، نسبه، فحص، العلاج، الشفاء، باذن، والمستشفيات، بادرى، السرطان، انواع، الوعي
Translation	Breast, Cancer, Early, Cancer, October, Month, Awareness, Examination, Campaign, Injury, Association, for awareness, to support, Addition, Tumors, Women, Symptoms, Protection, to detect, Percentage, Checkup, Treatment, Healing, with permission, Hospitals, Hurry up, Cancer, Types, Awareness

"تكمن أهمية الكشف المبكر لمرضى السرطان في زيادة نسبة الشفاء بإذن الله."

"The importance of early detection of cancer patients is increasing the rate of recovery, God willing."

Another tweet was posted by a news account, and indicates that 50 types of cancer can be detected through a blood test.

"فحص دم ثوري يمكنه كشف 50 نوعاً من #السرطان قبل ظهور الأعراض."

"A revolutionary blood test that can detect 50 types of #Cancer before symptoms appear."

The second service is Prevention and Control. Preventing and controlling the acceleration of cancer growth is an important key for preventing cancer from getting worse. The tweets related to this cluster involve important factors in cancer prevention, control, and survival such as lifestyle, behaviors, early diagnosis, and detection. The tweets associated with this topic were posted by doctors, patients, and others. For example, the following tweet was posted by a medical account. It highlights the importance of a healthy lifestyle in cancer control and prevention.

"يحدث نمط الحياة فرقا هائلا عندما يتعلق الأمر بالوقاية من مرض السرطان، حيث تعد التمارين الرياضية الجزء الأكثر أهمية في ذلك."

"Lifestyle makes a huge difference when it comes to cancer prevention, and the exercise is the most important part."

The third service is Causes. The keyterms represent the service including; Because, Exposed, Genetic, Mutation, and Radiation. Avoiding cancer causes is an essential strategy for preventing cancer development. Cancer causes were heavily discussed on Twitter by doctors, patients, families of patients, and others. The discussion from nonmedical experts was based on their common knowledge as well as their experiences about the causes of cancer. Genetic mutation could be one of the causes of cancer, which is inherited in some families. Other possible causes of cancer that includes exposure to chemicals, radiation, and sun rays were detected from the tweets. Sadness, the negative feelings, and the unhealthy lifestyle, and others were also mentioned in the tweets (See Section 5.1 for more details). For example, the following tweet, found in our dataset, was posted by a doctor.

"ماهي مسببات #السرطان؟ الجواب: عوامل متعددة، أبرزها: السمنة، التدخين والخمر، الوراثة، أعراض جانبية للأدوية، حالة الالتهاب."

"What are the causes of #Cancer? Answer: Multiple factors, most notably: obesity, smoking and alcohol, genetics, side effects of medications, and inflammation."

Furthermore, the following tweet highlights the risk of sun rays as one of the causes of cancer.

"التعرض الكثير لأشعة الشمس الضارة يسبب سرطان الجلد ،الله يعافينا وإياكم."

"Too much exposure to the harmful sun rays causes skin cancer, may God protect us and you"

The fourth service is Screening (see Row 5, Table 3) represented by keywords such as; Examination, Self-Examination, Mammogram, Doctor, Device, Minutes, Clinic, and others. Screening tests focus on detecting disease before symptoms appear. Early detection of cancer is very crucial for preventing the disease from reaching late stages, reducing complications, facilitating treatment, making treatment more successful, and consequently, reducing the mortality risks. Thus, screening tests are very significant; Prevention. Different screening tests are used to detect cancer diseases such as mammogram, clinical breast examination, and breast self-examination for breast cancer, and pap smears, and human papillomavirus test for cervical cancer. The mammogram screening test is an X-ray image of the breast. Most of the tweets, which are related to this cluster, stress the importance of the examination. The following tweet was posted under the hashtag #breast_cancer on 26 October 2021, by the government.

"#سرطان الثدي أكثر أنواع أمراض السرطان شيوعا لدى النساء، وننصح بإجراء الفحوص الدورية بجهاز "الماموغرام" الذي يساعد في الحد من حالات الوفاة من المرض."

"#Breast_Cancer is the most common type of cancer in women, and we recommend regular examinations with a "mammogram" device, which helps reduce death from the disease."

Another tweet in our dataset is posted by a medical account. It indicates the possibility of detecting cancer before it occurs.

"......... يمكن كشف سرطان الثدي قبل حصوله ب ٣ سنوات، أشعة الماموغرام للثدي تكشف السرطان قبل حصوله بثلاث سنوات."

" . . . . . . Breast cancer can be detected 3 years before it occurs, and a mammogram detects cancer 3 years before it occurs."

The fifth service is Symptoms. Observing symptoms is a vital factor for early treatment and preventing the acceleration of the disease. This cluster relates to the symptoms associated with cancer. Cancer symptoms vary from case to case depending on the organ affected by cancer. Symptoms of cancer include fever, sweating, constipation, loss of appetite, bleeding, changes in skin color, loss of body weight, and lumps or swelling under the skin. The tweets related to this cluster were posted by doctors, patients, and other stakeholders. For example, the following tweet, obtained from our dataset, was posted by a consultant oncologist about ovarian cancer symptoms.

"سرطان المبيض هو أحد أنواع السرطانات ......أعراضه: الم او تورم أو شعور بالضغط في منطقة البطن والحوض، نزيف من المهبل في غير وقت الدورة، افرازات من المهبل قد تُصحب بدم، انتفاخات أو امساك."

"Ovarian cancer is a type of cancer … Its symptoms: pain, swelling or a feeling of pressure in the abdomen and pelvis area, bleeding from the vagina outside the time of the period, secretions from the vagina that may be accompanied by blood, and swelling or constipation."

Moreover, we found another tweet related to cancer symptoms in an advanced stage. The tweet was posted by an account of a department in a hospital.

" في مراحل متقدمة، عندما ينتشر السرطان ويصبح ورم خبيث، قد تظهر اعراض مثل: اصفرار في الجلد، الم في العظام، صداع، مشاكل في التنفس."

"In advanced stages, when the cancer spreads, and becomes a malignant tumor, some symptoms may appear such as: yellowing of the skin, pain in the bones, headache, breathing problems."

Another tweet explained one of the symptoms of cancer, which is sudden weight loss.

" انا لاحظت نزول الوزن عندي بدون سبب، وبعد ما خذت منظار، طلع عندي ورم بالقولون ...."

"I noticed that I lost weight for no reason, and after colonoscopy, I had colon cancer … … … "

4.3. Treatment

We now discuss the services related to the second macro-service, Treatment. It includes Chemo and Radiation Therapy and Surgical Therapy. The sixth service is Chemo and Radiation Therapy (see Table 3). The keywords represent the service, for example; Chemotherapy, Pain, Duration, Receive, and Use. Chemotherapy and Radiation are among the important treatment types for cancer diseases. The chemotherapy is usually used to kill cancer cells and is given via a vein or by mouth. Radiation is used directly with a tumor through using high doses of radiation. These types of treatments help to destroy the tumor and improve the patient's clinical condition, prevent the spread of the tumor, and stop or slow the growth of the tumor. The main difference between chemotherapy and radiation is that chemotherapy takes medical drugs that target the whole body, while radiation therapy targets cancer cells in specific areas of the body. The tweets under this cluster reflect doctors' experiences, patients' experiences and concerns about treatment option, duration, and associated pain. The tweets related to this topic were posted by official medical accounts, patients, and other stakeholders. Some tweets provide information, for example, the following tweet, found in our dataset, was posted by a doctor.

"الخيارات العلاجية كثيره في حال اكتشاف سرطان الثدي :الجراحة "استئصال الثدي او الورم + فحص الغدد اللمفاويه او استئصالها"، العلاج الاشعاعي، الكيماوي العلاجات الهرمونية، المناعية، الموجهة. تعتمد على معطيات كل حاله على حده وعلى تفاصيل العينة بشكل كبير."

"There are many treatment options if breast cancer is detected: Surgery "mastectomy + examination or removal of lymph nodes", radiotherapy, chemotherapy, hormonal, immunomodulatory, and targeted therapies. It depends on the data of each case individually and on the details of the sample."

Other tweets reflect the patients' and stakeholders' experiences about the treatment (their concerns, treatment duration, and associated pain). For instance, the following tweet was posted by one of the stakeholders.

"ابنه صديقتي ذات الخمسة أشهر.. قرروا استئصال عينها بسبب ورم على الشبكية والعلاج الكيماوي للعين الأخرى."

"My friend's five-month-old daughter, they decided to remove her eye because of a tumor on the retina and chemotherapy for the other eye."

The following tweet was posted by a patient. It shows the patient concern about the chemotherapy.

"باقيلي 4 ايام على بداية الاسبوع الثاني من العلاج الكيماوي الشهر الأول بيكون مكثف، ......، أتمنى جسمي يتقبل العلاج ....."

"I have 4 days left until the beginning of the second week of chemotherapy, the first month will be intense, . . . , I hope my body accepts the treatment . . . . . . "

The seventh service is Surgical Therapy. It is characterized by keywords such as Malignant, Resection, Removal, Operation, Surgery, Tumor, and others. Surgical operation is one of the traditional cancer treatments. It is considered very effective in killing most types of cancers before the disease spreads to lymph nodes or distant sites (metastasis). Surgical treatment may be used alone or in combination with other treatment modalities, such as radiation therapy and chemotherapy. This option is taken if the cancer does not metastasize. During the surgery, doctors often remove lymph nodes near the tumor to see if cancer has spread to them. The tweets associated with this topic were posted by stakeholders' accounts, news accounts, patients, and other stakeholders. The tweets belonging to this topic include tweets reporting cancer surgeries, peoples' experiences with surgeries, and tweets asking for a financial or a moral support. The reported surgeries include those that have been completed and the ones that would be performed in the future. For example, the following tweet, obtained from our dataset, was posted on 14 September 2021, by one of the Saudi news accounts on Twitter. The tweet mentions the success of a tumor removal surgery that happened at a hospital in Taif city. The tweet was also announced by an electronic newspapers [59].

" استئصال "ورم سرطاني" زنة 7كجم من بطن سيدة .... بالطائف. "

"Resection of a "cancerous tumor" weighing 7 kg from the abdomen of a woman . . . . in Taif."

Furthermore, we found another tweet posted by a cancer patient. It highlights the need for moral support.

" متابعيني الكرام..... سوف اجري عملية استئصال ورم....... دعواتكم. "

"Dear followers . . . .I will perform a tumor removal surgery at . . . . .pray for me."

4.4. Psychological Support

We now discuss the services related to the third macro-service, Psychological Support. Cancer diseases affect the patients on a physical, spiritual, and emotional level. Therefore, psychological support is very important for cancer patients. It helps the patients to handle the difficulties and overcome challenges, which is an essential factor for the treatment success and survival. This macro-service includes three services; Spiritual Support, Suffering from Cancer, and Hope and Optimism. The eighth service is Spiritual Support (see Row 9, Table 3). Spiritual Support is one of the methods for providing psychological support. A common and important way for providing Spiritual Support is prayers (supplications) for cure, recovery, and patience. Making prayers are very important element in Muslims' beliefs and therefore, the patients, their families, friends, and beloved ones increase in their prayers. The keywords for this topic such as God, Cancer, Patients, Heal, Bodies, Strength, and Muslim represent the label. A lot of tweets, found in our dataset, were similar to the following tweet.

"دعواتكم لوالدي مريض سرطان."

"Pray for my father who has cancer."

"اللهم اشفي مرضى السرطان وبرد عليهم جرعات الكيماوي.."

"Oh God, heal cancer patients and make the chemotherapy doses easy on them . . . "

"اللهم اشف كل جسد أرهقه مرض السرطان."

"Oh God, heal cancer patients, they are in pain."

The ninth service is Suffering from Cancer. This service represents the pain and difficulties faced by cancer patients. Providing psychological support (e.g., listening to the patient, holding hands, making prayers) is one of the strategies for helping patients fighting cancer. This service included keyterms that represent expressions of physical and psychological pain associated with cancer. For example, people have called cancer a "silent killer" since it could kill the body and reach an advanced stage without obvious symptoms. The tweets related to this service were posted by patients, their families, and their beloved ones, such as the following tweet, which describes the suffering caused by cancer disease.

"وصف احد الأطباء معاناه مرضى السرطان قائلا كذئب جائع ينهش في لحم وعظم انسان."

"One of the doctors described the suffering of cancer patients. He said, cancer is like a hungry wolf that eats the meat and bone of a cancer patient."

Another tweet was posted by a patient. It highlights the physical pain associated with treatment.

" .. الكيماوي أصعب من المرض نفسه، مهما اتكلمت عمري محا اعرف اعبر عن كمية الالم اللي الكيماوي بيسويه ......
حاجه جواك تحرقك وتحرق روحك لدرجة حتى ملابسك ما تتحملها من الحرارة ما في مسكنات ولا ادويه ...."

" . . . chemotherapy is more difficult than the disease itself, whatever I talk I will not be able to express how much pain that chemotherapy causes . . . . . . something inside you burns you and your soul to the point where even your clothes can't stand the heat. There are no painkillers or medicines . . . ."

The tenth service is Hope and Optimism. Hope and optimism are among the most important factors that make a person live a happy and a healthy life. Many people are exposed to major psychological pressures and crises that may affect their lives, such as having the chronic cancer disease. A healthy psychological state is essential for the success of any treatment and recovery, therefore, providing psychological support is an important part of treatment. Hope, optimism, and patience are among the components of a healthy psychological condition. Helping cancer patients to be patient, willing, optimistic, and hopeful can be achieved through conversations, sharing experiences, and positive thoughts. Twitter is one of the social platforms where people can share their stories and experiences, influence and inspire other people. Some tweets were posted by patients who are fighting the disease, spreading hope, inspiring other people, and being a source of optimism. The tweets related to this topic were posted by patients, their families, and other stakeholders. For example, the following tweet found in our dataset is related to this cluster.

" .. مجرد الاستماع إلى قصته خلال مراحل محاربته للمرض تجعلك تعيد النظر في الكثير من
الأمور في حياتك لترى الحياة بمنظور أخر.."

" . . . just listening to his story during the stages of his fight against the disease makes you reconsider many things in your life to see life from another perspective . . . .."

Some tweets were posted by people who survived cancer and shared their successful experiences. The following excerpt of a tweet is an example of this cluster.

" .... هزمت السرطان 4 مرات. ومجبر على أن أكون قويا ...."

" . . . I defeated cancer four times and I'm forced to be strong . . . "

4.5. Socioeconomic Sustainability

Now we discuss the fourth macro-service Socioeconomic Sustainability. It involves the tweets and clusters that are related to developing better strategies to fight cancer in the country. It includes four services; the first service (the eleventh overall) is Government Support. The Saudi government is making a lot of efforts to confront cancer disease. These are represented in the provided medical services e.g., (healthcare centers, modern medical devices, free diagnosis and treatment, awareness, psychological support, call centers, and smart applications for healthcare systems). For example, the Ministry of Health has provided various services e.g., "Mawid" website and smart application for facilitating booking medical appointments, and a call center service that is available 24/7 on (telephone number 937) for normal and emergency health services to the patients across the Kingdom. "Shefaa Platform" is another example of the health services supported by the government. It facilitates the provision of medical services, medical devices, and medicines to the needy and the emergency cases for those who cannot obtain treatment in health facilities. It contributes to the governance of charitable treatment in the Kingdom by verifying the Absher platform and the Council of Health Insurance. People used Twitter to facilitate communication with the community. For example, some individuals use Twitter to accelerate receiving personal support (financial, treatment, or other). They usually attach links (e.g., for payment) from reliable platforms such as Shefaa. For example, the following tweet, was found in our dataset.

"سيدة .. تعاني من تدهور حالتها الصحية اثر وجود ورم سرطاني في الثدي .. مما يستدعي ضرورة التدخل
العاجل بالعلاج الكيماوي اللازم قبل وصولها لمرحلة حرجة يصعب فيها التحكم بالورم."

" … a woman suffers from a deteriorating health condition due to the presence of a cancerous tumor in the breast … which calls for urgent intervention with the necessary chemotherapy before it reaches a critical stage in which the tumor is difficult to control."

The government, individuals, healthcare institutions, and other stakeholders also used Twitter to announce new services, and share information and experiences on available services. This cluster represents the tweets related to government support and the healthcare services in Saudi Arabia. For example, the following tweet was posted by a doctor. It highlights some of the services provided by the Ministry of Health.

"سرطان الثدي يأتي بالترتيب الأول بين أكثر انواع السرطان شيوعا، عالميا، وإقليميا، ومحليا، لا يوجد اعراض
بالمراحل المبكرة له ف من المهم الكشف المبكر عن طريق الماموغرام، للحجز عن طريق الرقم 937 او تطبيق
" موعد " . مصدر المعلومات: وزارة الصحة." #سرطان_الثدي

"Breast cancer ranks first among the most common types of cancer, globally, regionally, and locally. There are no symptoms in its early stages, so it is important to detect it early through a mammogram. For reservations, call 937 or the "Mawd" application. Source of information: Ministry of Health." #Breast_Cancer

The twelfth service is Socioeconomic and Operational Challenges (see Row 13, Table 3). It is characterized by the keyterms such as Patient, Treatment, Needs, To Receive, Ask, Surgery, Thousand, Bill, Case, Suffering, Number, and Please. This cluster represents the socioeconomic and operational challenges faced by patients and their families. For example, one of the challenges is the need for patients and their families to travel farther to specialized hospitals for diagnosis and treatment purposes. This involves difficulties from various perspectives. For example, from the financial perspective, it involves additional expenses for transportation and housing. Another challenge is blood shortage. Some patients need a frequent blood transfusion and sometimes it is difficult to find donors, especially in urgent cases. The following excerpt of a tweet is an example of this cluster.

"اختي مصابه بـ لوكيميا " سرطان الدم " . اللي عنده استطاعة محتاجين متبرعين بالدم بشكل عاجل،الفصيلة: O+ ،
المكان: مستشفى الملك خالد الجامعي، رقم الملف: .....، الاسم: ......"

"My sister has leukemia. Those who can, she needs blood donors urgently, blood group: O+, location: King Khalid University Hospital, File No.: … , Name: … "

The thirteenth service is Charity Organizations. There are various charity organizations in the country, for example, Zahra Association, Sanad Children's Cancer Support Association, the Saudi Cancer Society, and The Civil Society Association for Cancer Care (Basma). These charity organizations are non-profit associations whose goal is to support cancer patients. They aim to provide social, financial, and psychological care and support for combating cancer. Their programs include, among others, encouraging new patients to contact other patients through conducting meetings and exchanging stories and experiences. For example, the Saudi Cancer Society holds workshops for patients and their families under the supervision of psychiatrists. These organizations may also participate in cancer awareness campaigns. For example, the following tweets shows that some organizations have participated and people are thanking and appreciating their efforts.

"كل الشكر موصول لجمعية ... على ما بذلته من جهود في حملة سرطان الثدي التي أقيمت بالأمس بتاريخ ١٧-١٠-٢٠٢١ ..."

"All thanks go to [organization] for its efforts in the breast cancer campaign that was held yesterday on 17-10-2021 . . . ."

The fourteenth service is Financial Support (Row 15, Table 3). It is represented by the keyterms such as; Treatment, Bill, Pinching, Tried, We Can't, and Appeal. This cluster represents the financial issues faced by cancer patients and their needs for financial support. The Saudi government provides free treatment services and cancer patients usually take this option. There is a waiting time (usually long queues) for this free treatment service. However, there are some cases where a cancer patient needs a quick treatment procedure (e.g., urgent operation) and cannot wait for the free service. Therefore, they go for private treatment, which is usually very costly. Some patients who are unable to afford the treatment costs use Twitter to raise their needs and request financial support.

"أب ... يعاني من ورم في البروستات ... سبب له آلام حادة ويحتاج إلى التدخل الجراحي بشكل عاجل، ولا يملك القدرة المادية على تحمل تكاليف العملية."

"A father . . . suffers from a prostate tumor . . . which caused him severe pain and needs urgent surgical intervention, and he does not have the financial ability to bear the costs of the surgery."

Furthermore, cancer disease can severely affect patients' financial state for some reasons such as inability to work or loss of job due to health status. These situations create a need for financial support for some cancer patients.

" .... اناشدكم ...مريضه سرطان ابغى سداد الفاتورة زوجي متوفى ... وذلك مصاريف علاجي من السرطان الراتب 3500 اجار شقه متراكم."

" . . . .I appeal to you, . . . a cancer patient, I want to pay the bill. My husband is deceased. . . . . This is a treatment expense for cancer. My salary is 3500 SR. Apartment rent is accumulating."

"المساهمة بالرتويت كمساهمتك في دفع الفاتورة، ام مصابه بالسرطان ....،المتبقي 32800، تكفون يا اهل الخير."

"Contribute by retweeting as your contribution to paying the bill. Mother has cancer . . . . . . , Remaining is 32,800 SR, please good people."

4.6. Information Availability

We now discuss the services associated with the fifth macro-service, Information Availability, which means that information is conveniently available to all stakeholders including patients, the families of the patients, healthcare organizations, and government. Information availability is a key strategy in combating cancer for all stakeholders. It includes three services. Breast Cancer Awareness (fifteenth service created from merging Topic 9 and 20) includes the keyterms; Campaign, Association, Center, Awareness, Lecture, and Examination. According to Breast Cancer Organization, breast cancer is the most common cancer across the globe [60]. It is the number one type of cancer among the women globally, as well as in Saudi Arabia [61]. Therefore, breast cancer has gained more attention than other kinds of cancer types and it is also evident from our detected clusters. The aim of breast cancer awareness programs is to raise public awareness of breast cancer, the importance of early detection, the causes, and the prevention. One of the strategies for disseminating awareness about breast cancer was launching the annual campaign, "Breast Cancer Awareness Month". In October 2021, many initiatives for supporting breast cancer awareness were discussed on Twitter under several hashtags such as, "#اكتوبر_الوردي" (pink October), "#الشهر_العالمي_لسرطان_الثدي" (breast cancer awareness month), and "#افحصي_الان" (check-up now). The campaign was supported by medical institutions (e.g., hospitals, and healthcare centers), charity organizations (e.g., Ehsan), and other stakeholders. The following is an excerpt from a tweet that was posted on Oct 3, 2021, by the Public Health in Taif city.

"تم اليوم، ...، تدشين حمله سرطان الثدي، والتي تستمر طيلة شهر أكتوبر."

"Today, . . . , the breast cancer campaign was started, which will continue throughout the month of October."

Another similar tweet highlights the activation of breast cancer awareness campaign in Sabya governorate in Jazan region.

"مركز صحي ... بالتعاون مع جمعيه ... تفعيل حمله شهر اكتوبر للتوعية بسرطان الثدي."

"[Health Center] in cooperation with [association] activating the October campaign to raise awareness of breast cancer."

The sixteenth service is Awareness Campaigns. It is characterized by keywords including; Account, Question, Retweet, Reply, Hashtag, Tweet, and Campaign. These key terms belong to Topic 10; see Figure 5. Table 5 gives the English translation of the Arabic keywords in the figure. Awareness campaigns are very important as they raise awareness of cancer types, causes, prevention, and treatment. Moreover, these campaigns can encourage early detection, support people with cancer, encourage cancer control programs, and enable communications and sharing of experiences between patients. Awareness campaigns could be conducted by any healthcare stakeholders though usually these campaigns are carried out by healthcare providers. The campaigns involve various activities for disseminating information such as lectures, posters, brochures, leaflets, and competitions. Awareness campaigns may also involve raising Twitter hashtags and sharing on Twitter posts containing educational content. For example, we detected various hashtags that were raised supporting cancer awareness campaigns including "#اليوم العالمي للسرطان" (world cancer day), "#سرطان البروستات" (Prostate cancer), "#سرطان الدم" (Blood Cancer), "#الشرقية_وردية_١٣" (Al-Sharqiya Pink 13), and others. The tweets related to cancer awareness campaigns were posted by doctors, hospitals, healthcare centers, and organizations. For example, the following tweet was posted by a hospital in Makkah.

"اليوم يبدأ #الشهر_العالمي_لسرطان_الثدي... ، الى كل بطلة تحارب
تذكري دائما أنك وبالإيمان والعزيمة."

"Today starts the #International_Breast_Cancer_Month . . . to every hero who fights, always remember that you are with faith and determination."

Furthermore, a tweet was posted under (Al-Sharqiya Pink 13) hashtag, which was launched for supporting an event under the breast cancer awareness campaign. The event was organized by the Saudi Cancer Society in the Al-Sharqiya region. It involved various activities including the event of "Sharqiya Pink Marathon". It aimed to support the global trend in raising awareness of the disease and the importance of early detection as one of the factors influencing the stages of treatment, which makes the difference in combating and curing this disease.

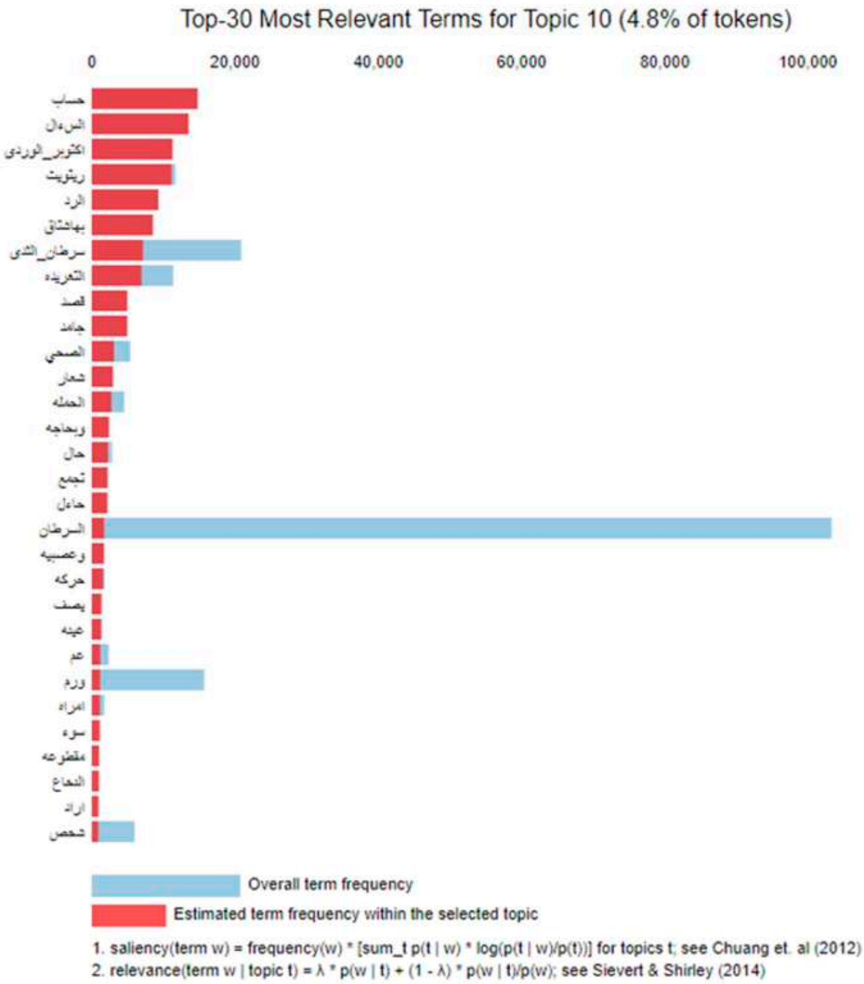


Figure 5. The Top 30 Most Relevant Terms for Topic 10.

Table 5. The keywords and their English translation for Figure 5.

Keywords	حساب، السؤال، اكتوبر_ الوردي، ريتويت، الرد، بهشتاق، سرطان_ الثدي، التغريده، قصد، جامد، الصحي، شعار، الحمله، وبحاجه، حال، تجمع، حاءل، السرطان، وعصبيه، حركه، يصف، عينه، عم، ورم، امراه، سوء، مقطوعه، النخاع، اراد
Translation	Account, The Question, Pink October, Retweet, Reply, Hashtag, Breast Cancer, Tweet, Intent, Solid, Healthy, Logo, The Campaign, Need, Case, Gathering, Hail city, Cancer, Nervous, Movement, Describe, A Sample, Uncle, Tumor, Woman, Bad, Cut Off, Marrow, Want

The seventeenth service is Questionnaires and Competitions (Row 19, Table 3). It is represented by keywords such as; Follow, Participation, Share, Event, Member, Answer, Thousands, and more. Information availability is not only about creating awareness for care receivers; it is about gaining knowledge from the care revivers and public for scientific research purposes and improving healthcare. One of the strategies to enhance societal awareness and educate people about cancer about its risks and the ways to prevent it is the development of certain questionnaires and competition exercises among the public. Questionnaires and competitions can also be used to obtain information for scientific research purposes. They can be organized and supported by various organizations, associations, and electronic platforms. For example, one of the charitable organizations for cancer patients, held various online competitions on Twitter and provided financial incentives for stimulating awareness among the members of the community. An example tweet is given below.

"#مسابقة بسمه /السؤال الثالث: ما هي أسباب الإصابة بسرطان الثدي؟........"

"#Basma competition/Question three: What are the causes of breast cancer? . . . ."

5. Extended Services

This section discusses the 42 extended services that we discovered by using knowledge gained from and extending the initial discovery process. These services are discovered by *topical searching* over the dataset, the *topics* are some of the services discovered in the initial discovery process. We used four *topical searches* (Causes, Symptoms, Prevention, and Stakeholders) and detect 42 additional services; these are discussed in Section 5.1 to Section 5.4, respectively. We call them extended services since these are discovered by using the knowledge gained from the initial service discovery process.

5.1. Cancer Causes

We intended to investigate the usability of our approach further and hence we applied clustering on our dataset with a focus on cancer causes. This approach allowed us to detect a list of cancer causes from the collected tweets. We used certain keyterms to find more about the causes. These terms are some important Arabic terms that refer to the causes of cancer for example "تسبب", "يسبب", "بسبب", "مسبب". We categorized the causes into several groups based on the content of the tweets. The causes of cancer that we have detected include Carcinogenic Foods, Wrong Eating Habits, Genetic Causes, Obesity, Smoking, and others. The full taxonomy for the causes of cancer that we extracted from Twitter data is presented in Figure 6.

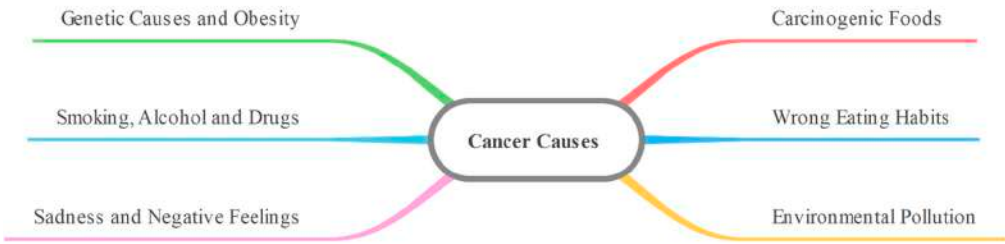


Figure 6. Cancer Causes Taxonomy Extracted from Twitter Data.

Carcinogenic Foods are foods that contain chemicals such as preservatives, taste enhancers, artificial flavors, and juices to which preservatives are added. Eating a lot of carcinogenic foods may cause cancer. It is advised to commit to a healthy diet that contains nutrients and avoid harmful foods. The following are sample tweets that elaborate on this. They were posted by patients, doctors, and other stakeholders. *" . . . .they found out that my nephew has leukemia, the reason is Indomie and soft drinks . . . ". "Be careful, juices and soft drinks cause colon cancer for young people and various chronic diseases. Get to know them and avoid them immediately."*

There are some Wrong Eating Habits that increase the risk of cancer. These include some cooking methods, for example, deep-frying, cooking food more than necessary at high temperatures, which generate chemicals like heterocyclic amines, and the use of plastic for food preservation. The following tweets were posted by patients and other stakeholders.

"لا تغل الطعام المطبوخ أكثر من اللازم: ينتج عن كثرة غليان هذا الطعام تولد مواد تنشط الخلايا السرطانية وتزيد من تكاثرها، لذا يكفي غليه مرة واحدة بعد طهيه، ويتم تناوله بعد طهيه مباشرة."

"Do not boil cooked food more than necessary: frequent boiling of this food generates substances that activate cancer cells and increase their proliferation, so it is sufficient to boil it once after cooking it, and eat it immediately after cooking."

Negative Feelings and emotions, such as anger, sadness, and hate, negatively affect and weaken the immune system. Therefore, it is necessary to control emotions to decrease the risk of getting cancer. The following tweets represent the cluster. They were posted by patients and other stakeholders.

"مشاعرك السلبية تتكدس في عضو معين من جسدك اذا لم تواجهها وتحررها وتتحول لا سمح الله الى مرض مزمن
..الأورام= جروح دفينة .السرطان = حزن عميييييق......"

"Your negative feelings accumulate in a certain part of your body, if you do not confront them and release them it may turn into a chronic disease. . . . . Tumors= buried wounds. Cancer = deep sadness . . . ."

Smoking, Alcohol, and Drugs are among the most important causes of cancer that negatively affect the health of the body. The following are sample tweets, which elaborate on that.

" ماهي مسببات #السرطان؟ الجواب: عوامل متعددة، أبرزها: السمنة، التدخين والخمر، الوراثة، أعراض جانبية
للأدوية، حالة الالتهاب."

"What are the causes of #cancer? Answer: Multiple factors, mostly: obesity, smoking and alcohol, genetics, side effects of medications, and inflammation."

Genetics is an important factor in diseases in general and cancer in particular. Some cancer diseases have a genetic predisposition, and some are generated from genetic mutations. Obesity is another risk factor for cancer. It changes hormones secretion, affects the immune system, and causes chronic cellular inflammation and these can lead to cancer. The following tweet is posted by a doctor's account, which elaborates on that, *"What are the causes of #cancer? Answer: Multiple factors, mostly: obesity, smoking and alcohol, genetics, side effects of medications, and inflammation."*

Environmental Pollution is one of the most important causes of cancer. It includes air pollution with toxic gases or factory smoke, and water pollution with bacteria and rust. In addition, exposure to harmful rays increases the chances of developing cancer. The following tweets represent the cluster. These tweets were posted by patients, the public, and other stakeholders.

"... مرض السرطان .... اغلب المدن تعاني منه .... من المصانع والمواد الغذائية الملونة."

" . . . cancer disease . . . most cities suffer from it . . . caused by factories and colored foodstuffs."

5.2. Cancer Symptoms

As part of our investigation of usability of our approach, we applied clustering to our dataset with an emphasis on cancer symptoms. From the collected tweets, we were able to derive a list of cancer symptoms. We used some search queries with certain keyterms. These keyterms refer to the cancer symptoms such as "اعراض", "الاعراض", "علامات". We obtained important cancer symptoms, for example; Fever, Fatigue, Bleeding, Discharge, Weight Loss, Headaches, and others. The full taxonomy for the symptoms of cancer that we extracted from Twitter data is presented in Figure 7. The tweets were posted by stakeholders of healthcare. For example, below is a tweet that shows the symptoms associated with ovarian cancer.

"سرطان المبيض هو أحد أنواع السرطانات .... أعراضه: الم او تورم او شعور بالضغط في منطقة البطن
والحوض، نزيف من المهبل في غير وقت الدورة، افرازات من المهبل قد تصحب بدم ، انتفاخات اوامساك."

"Ovarian cancer is one of cancer diseases . . . . its symptoms: pain, swelling or a feeling of pressure in the abdomen and pelvis area, bleeding from the vagina outside the time of the period, secretions from the vagina that may be accompanied by blood, and swelling or constipation."

The following tweet highlights breast cancer symptoms such as Swelling or Lumps and the change in Size of the Breast: "Another clear symptom of breast cancer is the emergence of swellings under one or both armpits, as a result of swelling of the lymph tissues there. A noticeable change in the size of the breast is unjustified, as it swells greatly. A noticeable shrinkage or retraction of the nipple inward, which is one of the important symptoms of breast cancer." Weight Loss symptom is shown in the following tweet. " . . . I noticed that I lost weight for no reason, and after I took an endoscopy, they found that I have a colon tumor . . . ". The following tweets shows some symptoms such as; Yellowing of the Skin, Pain in the Bones, Headaches, and Breathing Problems. "In advanced stages, when the cancer spreads, and becomes a malignant tumor, some symptoms may appear such as: yellowing of the skin, pain in the bones, headache, and breathing problems."

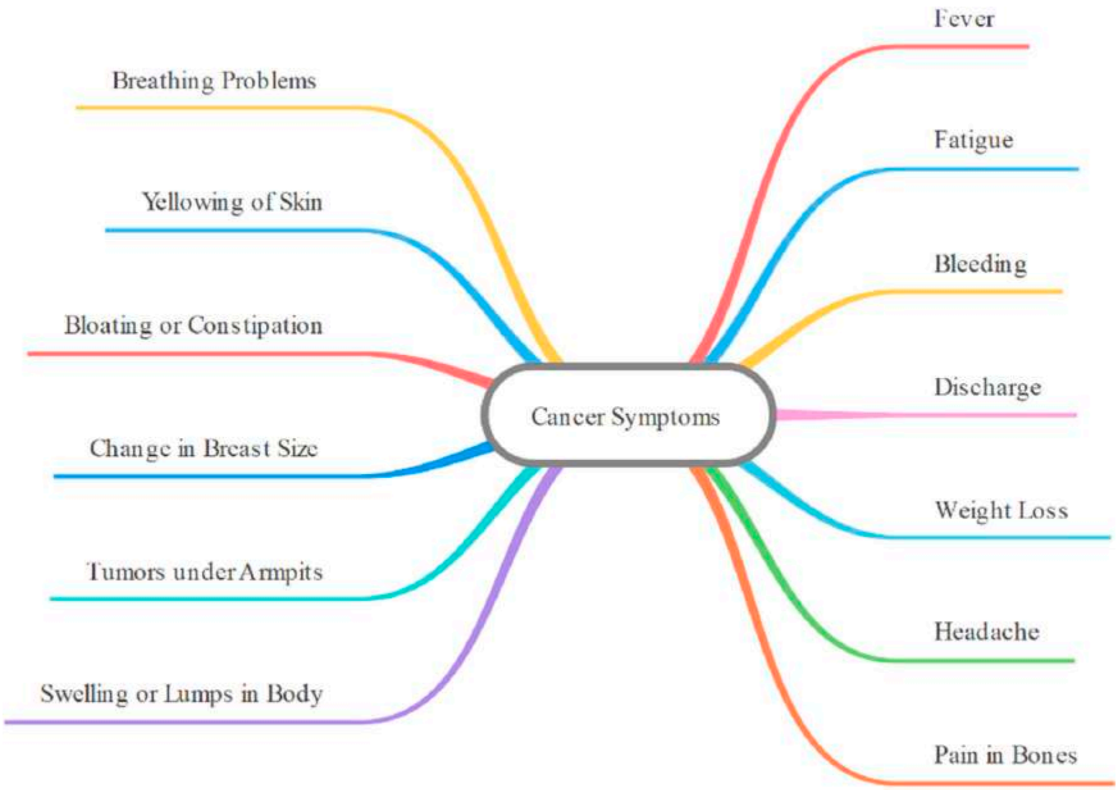


Figure 7. Cancer Symptoms Taxonomy Extracted from Twitter Data.

5.3. Cancer Prevention

We applied clustering on our dataset with a focus on cancer prevention. By analyzing the collected tweets, we were able to identify a list of cancer prevention methods. We used certain terms for instance "الوقاية", "وقاية", "للوقاية", "يقيك". We identified important ways to prevent cancer, which include; Sports, Healthy Diet, Healthy Lifestyle, No Smoking, etc. The full taxonomy that we extracted from Twitter data is presented in Figure 8. The following tweets were posted by the different accounts as patients, hospitals, doctors, and other stakeholders. For example, the following tweets, found in our dataset, are related to; Sports, Healthy Lifestyle, No Smoking, Alcohol Control, Vaccination, Avoiding Stress, Drinking Water, and Sleeping Early. *"Healthy nutrition, sports, drinking water, sleeping early, staying away from stress . . .* etc. *are all important factors for the prevention of all diseases and cancers in particular. Let your health always be your priority for a safe future from diseases, God willing." #Pink_October, #Breast_Cancer_Awareness. "Can you prevent cancer? Researches showed that up to 50% of cancer cases can be prevented through healthy lifestyle. You make choices every day that affect your health. Prevention and early detection are more important than ever: don't use tobacco, eat a healthy diet, be active and maintain a healthy weight, refrain from drinking alcohol,* *avoid risky behaviors, vaccination (human papillomavirus and hepatitis), know your family medical history and get regular checkups.”* Furthermore, the following tweet highlights Smoking and cancer. *“One of the simplest ways to prevent cancer is to stop smoking.”*

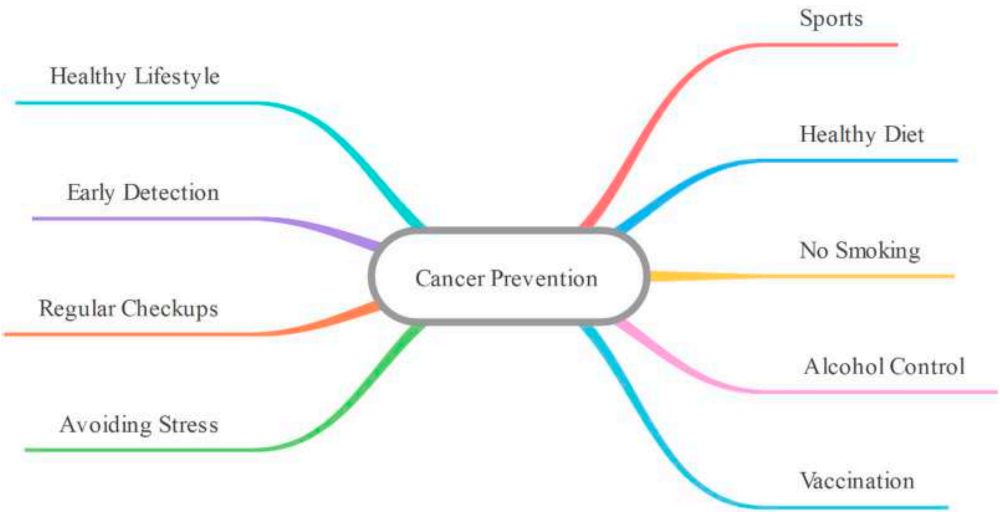


Figure 8. Cancer Prevention Taxonomy Extracted from Twitter Data.

5.4. Healthcare Stakeholders

We extracted a cluster for stakeholders in healthcare. We used some search queries with specific keyterms, such as “دكتور”, “وزارة الصحة”, “اصدقائي”, “عائلتي”, “مريضه”, “مريض”, “صيدلية”, “تحاليل”, “مختبر”, “عيادة”, “مركز صحي”, “مستوصف”, “مستشفى”, “ممرضه”, “ممرض”, “جمعية”, “تأمين”, “علاج”. A stakeholder is an individual or a group with an interest in the subject. In the healthcare sector, individuals or organizations with an interest in healthcare decisions are referred to as stakeholders. The stakeholders that we have detected include; Patients, Family, Friends, and others. The full taxonomy that we extracted from Twitter data is presented in Figure 9.

The following tweet represents the stakeholder, Patients: *“I am a cancer patient . . . ”*. The following tweet shows the Family and Friends stakeholders: *“I lived my experience with cancer with full strength, . . . Everyone around me shared my pain. I found constant support from my husband, my family, and my friends . . . ”* Below are some tweets related to the Ministry of Health stakeholder: *“The Ministry of Health invites people with chronic diseases and weak immunity and those who receive immunosuppressive drugs such as . . . cancer patients . . . to quickly go to the vaccination centers to receive the vaccine in order to avoid any complications that may threaten their lives if they get the virus.”* The following tweet represents the Doctors stakeholder. *“ . . . Doctor . . . Can you advise me about the Zometa injection? I am a breast cancer patient . . . ”* Hospitals as a stakeholder are shown in the following tweets: *“[Hospital . . .] succeeds in applying advanced technology to treat eye cancer in children*”.

We had collected one or more tweets for each stakeholder type and added them to the paper, however removed from the final version for brevity. We will be happy to provide the reader with samples of additional tweets if required.

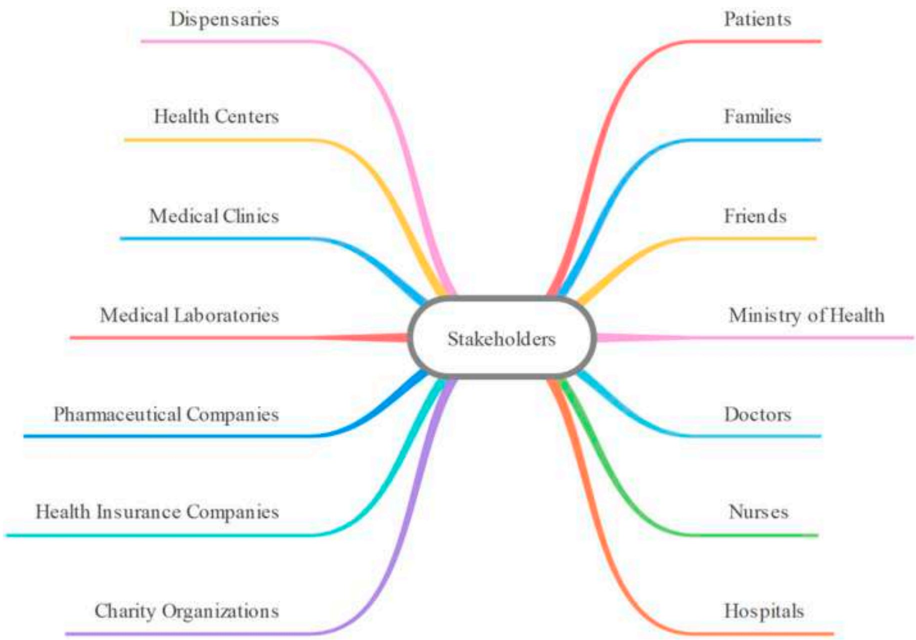


Figure 9. Taxonomy for Stakeholders in the Healthcare Sector Extracted from Twitter Data.

6. Discussion

In this paper, we proposed a data-driven artificial intelligence (AI) based approach called Musawah to automatically detect and identify healthcare services that can be developed or co-created by various stakeholders using social media analysis. We detected 17 services using unsupervised machine learning from Twitter data using the Latent Dirichlet Allocation algorithm (LDA) and group them into five macro-services.

The first macro-service, Prevention, involves actions that reduce the chance of getting cancer (e.g., maintaining a healthy lifestyle, taking vaccines, and avoiding cancer-causing substances) and it also involves strategies and procedures that prevent cancer from progressing to advanced stages (e.g., early detection, and regular screening tests). The services in this category include; Early Diagnosis, Prevention and Control, Causes, Screening, and Symptoms. The second macro-service, Treatment, describes the cancer treatment services and increases the level of understanding about options for cancer treatment, which can be used to treat cancer in different ways depending on the patient's condition and stage of the disease. The service in this macro-service includes Chemo and Radiation Therapy and Surgical Therapy.

The third macro-service is Psychological Support. It includes services, namely, Spiritual Support, Suffering, Hope and Optimism. Cancer diseases affect the patients on a physical, spiritual, and emotional level. Therefore, psychological support is very important for cancer patients. It helps the patients to handle the difficulties and overcome challenges, which is an essential factor for a successful treatment and patient's survival, even the survival of the patient's families. The fourth macro-service (Socioeconomic Sustainability) involves the tweets and clusters that are related to developing better strategies to fight cancer in the country. The services in this category include Government Support, Socioeconomic and Operational Challenges, Charity Organizations, Financial Support. The fifth macro-service, Information Availability, which means that information is conveniently available to all stakeholders including patients, the families of the patients, healthcare organizations, and government. Information availability is a key strategy in combating cancer for all stakeholders. It includes the; Breast Cancer Awareness, Awareness Campaigns, and Questionnaires and Competitions services.

We also introduced in this paper our methodology to develop extended services, called so as these are discovered by using the knowledge gained from the initial service discovery process. We discovered 42 extended services by topical searching over the dataset, where the topics are selected from some of the services discovered in the initial discovery process. We used four topical searches in this paper—Causes, Symptoms, Prevention, and Stakeholders—however, any number of services can be used to create services in a specific subdomain. The 42 services are shown in the extended taxonomy of cancer provided in Figure 10.

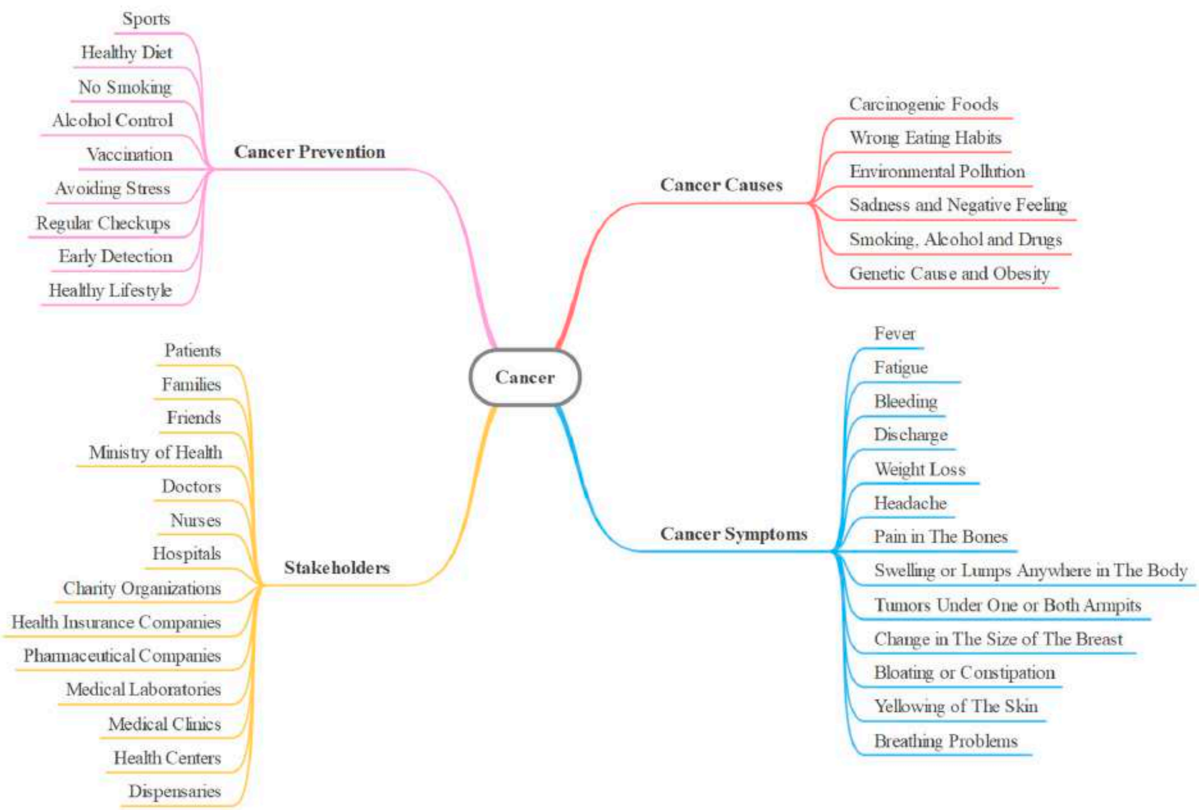


Figure 10. Extended Cancer Taxonomy: A Taxonomy of Extended Cancer Services.

We mentioned in Section 1 that the aim of this paper is to investigate the role of big data analytics over social media to automatically detect needs and value propositions that could be nurtured and turned into service co-creation processes through engaged participation over social and digital media, eventually co-creating services. The co-created services could be based on values that are not necessarily materialistic but are driven by equity, altruism, community strengthening, innovation, and social cohesion. We provide below evidence that big data analytics over social media can be used to detect value propositions and co-create values through engaged participation. Specifically, we provide a few examples below to elaborate the concept of automatic discovery, development, deployment of co-created services that we introduced in this paper. We take the example of the discovered service 'Hope and Optimism' that was discussed in Section 4. Hope and optimism are among the most important factors that make a person live a happy and healthy life. Many people are exposed to major psychological pressures and crises that may affect their lives, such as having chronic cancer disease. A healthy psychological state is essential for the success of any treatment and recovery, therefore, providing psychological support is an important part of treatment. Hope, optimism, and patience are among the components of a healthy psychological condition. Helping cancer patients to be patient, willing, optimistic, and hopeful can be through conversations, sharing experiences, and positive thoughts. People share their stories and experiences of pain and suffering, joy and happiness, struggles and overcoming diseases, stories of inspiration, optimism, and more. As we have mentioned earlier, service co-creation is enabled through engaged participation of need and value propositions. In co-creation terminology, both parties, a customer and company create value for each other and this value could be money, product, service, or anything of value for the parties involved. Therefore, the value could be anything materialistic or otherwise, such as a desire for equity, the altruistic nature of a person or party, and the desire to strengthen communities, bring innovation, or social cohesion. Now imagine an open service and value healthcare system based on freely available information made available by the public, the government, third and fourth sectors, or others. Imagine someone in need of Hope and Optimism shares the need on social media and this is matched by a value proposition, someone with a specific experience, knowledge, skillset, etc. These two people or parties come together to exchange value for value such as the need for hope and optimism exchanged for a monetary value or desire for equity and social cohesion, and in the process, they co-create services that can be used as it is in many other cases or adapted for different situations.

Another example is Suffering, which can be looked at as a service where the person who suffers has a need and another person or party can provide value to the one suffering directly (patient) or indirectly (patient's family) from cancer. Suffering due to cancer can be physical, psychological, or financial, and these sufferings are well known. For example, peoples' finances can be negatively affected due to cancer treatment for themselves or their family members and cause a person to suffer. Suffering can also be due to operational reasons, such as the hardships related to the need for patients and their families to travel farther to specialized hospitals for diagnosis and treatment purposes. This involves difficulties from various perspectives. For example, from the financial perspective, it involves additional expenses for transportation and housing. Another suffering could be due to blood shortage. Some patients need a frequent blood transfusion and sometimes it is difficult to find donors, especially in urgent cases. Using the previous example of Hope and Optimism service, in the case of Suffering, two people or parties may come together to exchange value for value, such as the need for relief from Suffering due to financial or operational reasons, or needing blood, exchanged for a monetary value or desire for equity and social cohesion, and in the process co-create a service.

Similarly, Information Availability could promote the value co-creation process for healthcare services by allowing healthcare providers to interact with patients, and patients with similar health situations for sharing information about conditions, symptoms, and treatments. This promotes collective learning and emotional support, helps individuals make decisions, spreads health knowledge, increases health literacy, reduces the patient's anxiety level, and provides personalized self-management. Socioeconomic Sustainability could relate to supporting healthcare services and value co-creation by using the available healthcare resources to reduce doctor and hospital visits and thereby reduce costs and wasted resources. Service co-creation could also relate to identifying the needs of a patient and meeting those needs to gain patient satisfaction by using feedback from the patient's or other patients' experiences acquired from Twitter data. The Treatment service could relate to increasing the level of understanding about options for cancer treatment available to a specific individual or group of people and these treatments may not be the mainstream methods rather based on well-known herbal medicine or lifestyle changes depending on the patient's condition and stage of the disease.

The Musawah approach and system proposed in this paper make important and significant theoretical and practical contributions to the literature. While there were a few works (e.g., [29]) on cancer-related social media studies in the English language, these were focused on one or another type of cancer such as breast or skin or lung cancer (see Section 2.2). There was only work on social media-related cancer studies and it was focused on studying chemotherapy misconceptions among Twitter users [41]. These works clearly are different from our work, as we have looked at the whole cancer space in this paper, not just a type of cancer or chemotherapy misconceptions. There are also some works, such as [54], that have investigated the use of social media by people and public health organizations to communicate in new ways for various mutually beneficial activities. While these and other works have investigated the use of social media for value co-creation by various actors communicating with each other and sharing information, none of the works in the literature have proposed to discover healthcare services automatically from social media. This work, therefore, contributes a novel approach of using social media and AI to discover, develop, and exchange healthcare or cancer services. We demonstrated the potential of our approach by discovering over fifty services in several dimensions of cancer disease including cancer causes, symptoms, prevention, treatment, socioeconomic sustainability, and stakeholders. The approach can be systemized by creating open service and value healthcare systems based on freely available information made available by the public, the government, third and fourth sectors, or others, allowing new forms of preventions, cures, treatments, and support structures to be discovered and exchanged between various parties.

The work provides evidence to support the general literature on data-driven smart cities research [1,2,62,63] and reinforces that policy and action on smart cities, healthcare, and other sectors should be supported with data and that social and digital media provides a convenient and important source of such data [64,65]. The topics detected by our system clearly show the possibility and benefits of our tool, allowing from social media public and stakeholder conversations the discovery and grasp of important dimensions of the cancer disease in Saudi Arabia (partly applicable internationally), such as public concerns, patient requirements, solutions for problems related directly and broadly to cancer such as cancer treatment, financial difficulties, psychological ordeals and traumas, operational challenges for the families of cancer patients and other information. The parameters and information learned through the tool can benefit the public in many ways, such as through being a source of information and allowing government and various institutions to improve their services, organizational approaches, foci, etc. The work is distinct from others in terms of the dataset, methodology and design, our innovative approach of using AI to discover services, and our findings.

As regards the potential impact of our work, we believe the impact could be foundational, colossal, and far-reaching. On a personal note, the authors casually discussed the findings of this research and mentioned it to their families, friends, and networks, and this unconscious activity generated a lot of impacts. Some people got motivated to change their lifestyle to become healthier, others decided to be careful in using cosmetics to minimize the cancer risk. An important impact was related to a close friend who found that he had some symptoms, which were mentioned in the paper and he decided to do some tests. To his surprise, the reason for the symptom he had was an infection that is a major cause of a certain kind of cancer. These impact examples are the result of casually mentioning the finding of our paper to a small network of authors. How about coordinated dissemination of these findings to larger networks and how about creating incubators for service co-creation using our proposed ideas and tool? Such efforts on a large scale, incorporated into data-driven digitally connected systems, where information is pushed to consumers based on certain environmental parameters (e.g., a person searching for some symptoms or their smart devices reporting a specific pattern) can lead to lifestyle motivations, early interventions, diagnosis, and ultimately reduction in diseases and healthy and sustainable societies.

7. Conclusions and Future Work

Poor healthcare systems affect individuals, societies, the planet, and economies. Smartization of cities and societies has the potential to unite individuals and nations towards sustainability and thereby also improve healthcare and other systems as it requires an engagement with our environments, an analysis of them, and for us to make sustainable decisions regulated by TBL. Healthcare systems are a major contributor to the deterioration of our planet's environment due to high energy usage, disposable supplies, etc. This environmental damage causes health problems that need to be medically treated while these treatments in turn further damage the environment, and thus a chain reaction occurs. There is a growing demand for the prevention of diseases and for reducing the need for disease monitoring, screening, and treatment to a minimum. This in turn requires an open space for innovations and dynamic interactions between healthcare stakeholders to understand and provide solutions, services, information and resource supply chains, and community support structures for timely interventions and disease prevention and progression.

To address these challenges, this paper proposed a data-driven artificial intelligence (AI) based approach called Musawah to automatically detect and identify healthcare services that can be developed or co-created by various stakeholders using social media analysis. The case study focused on cancer disease in Saudi Arabia using Twitter data analytics in the Arabic language. Specifically, we discovered 17 services using unsupervised machine learning from Twitter data using the Latent Dirichlet Allocation algorithm (LDA) and grouped them into five macro-services; Prevention, Treatment, Psychological Support, Socioeconomic Sustainability, and Information Availability. We also illustrated the possibility to find additional services by topical searching over the dataset using four topical searches (Causes, Symptoms, Prevention, and Stakeholders) and detected 42 additional services. We developed a software tool from scratch for this work. The tool implements a complete machine learning pipeline. The dataset we used contains over 1.3 million tweets collected from September to November 2021.

Our work makes several contributions to the literature including the Musawah approach for creating services proposed in this paper and the developed tool and techniques to detect cancer-related services. The work builds on our extensive research in social media analysis in English and other languages on topics in different sectors. The paper has focused on cancer-related services from Twitter data in Arabic, however, the proposed approach broadly can be used for any disease or purpose and in any language.

The idea of developing cancer-related services from social media analysis is motivated by the well-known science that social media facilitates value co-creation through stakeholder interactions allowing value creation or creation of new services. Therefore, the ideas of services and co-creation proposed in this paper per se are not novel, and the novelty lies in the proposed approach of systematically and dynamically developing an ecosystem of services for a particular disease that can be co-created on-the-fly by communities of stakeholders interacting over social media using automatic value and service detection. The services can be the needs of people seeking prevention or treatment such that a gap in the market can be detected for stakeholders to develop services for economic, social, or other types of remuneration and rewards. The approach that we present, while basic at this stage, has great potential and will allow further investigation and development of novel and innovative ways of developing healthcare services and systems, is green in terms of environmental, social, and economic sustainability, with lower costs allowing lifestyle changes, self-directed disease management, and managed care.

We believe systemizing this approach could lead to a revolution in sustainable healthcare, driven by communities co-creating social values with the exchange of services for services that are not necessarily materialistic, but driven by equity, altruism, community strengthening, innovation, and social cohesion. Open service and value healthcare systems based on freely available information can revolutionize healthcare in manners similar to the open-source revolution by using information made available by the public, the government, third and fourth sectors, or others, allowing new forms of preventions, cures, treatments, and support structures.

Author Contributions: Conceptualization, N.A., S.A. and R.M.; methodology, N.A., S.A. and R.M.; software, N.A. and S.A.; validation, N.A., S.A. and R.M.; formal analysis, N.A., S.A. and R.M.; investigation, N.A., S.A., R.M., A.A. (Ahmed Alzahrani), A.A. (Aiiad Albeshri) and I.K.; resources, R.M., I.K., A.A. (Aiiad Albeshri) and A.A. (Ahmed Alzahrani); data curation, N.A. and S.A.; writing—original draft preparation, N.A., S.A. and R.M.; writing—review and editing, R.M., A.A. (Ahmed Alzahrani), A.A. (Aiiad Albeshri) and I.K.; visualization, N.A. and S.A.; supervision, R.M. and A.A. (Ahmed Alzahrani); project administration, R.M., A.A. (Aiiad Albeshri) and I.K.; funding acquisition, R.M., A.A. (Aiiad Albeshri) and I.K. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge with thanks the technical and financial support from the Deanship of Scientific Research (DSR) at the King Abdulaziz University (KAU), Jeddah, Saudi Arabia, under Grant No. RG-6-611-40. The experiments reported in this paper were performed on the Aziz supercomputer at KAU.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data is acquired from Twitter under their terms and conditions. Anyone can request the same data from Twitter.

Acknowledgments: The work carried out in this paper is supported by the HPC Center at the King Abdulaziz University. The training and software development work reported in this paper was carried out on the Aziz supercomputer.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yigitcanlar, T.; Butler, L.; Windle, E.; DeSouza, K.C.; Mehmood, R.; Corchado, J.M. Can Building "Artificially Intelligent Cities" Safeguard Humanity from Natural Disasters, Pandemics, and Other Catastrophes? An Urban Scholar's Perspective. *Sensors* **2020**, *20*, 2988. [CrossRef] [PubMed]
2. Mehmood, R.; See, S.; Katib, I.; Chlamtac, I. *Smart Infrastructure and Applications: Foundations for Smarter Cities and Societies*; Springer International Publishing, Springer Nature Switzerland AG: Cham, Switzerland, 2020.
3. Komeily, A.; Srinivasan, R.S. Sustainability in Smart Cities: Balancing Social, Economic, Environmental, and Institutional Aspects of Urban Life. In *Smart Cities: Foundations, Principles, and Applications*; Wiley: Hoboken, NJ, USA, 2017; pp. 503–534.
4. Alotaibi, S.; Mehmood, R.; Katib, I.; Rana, O.; Albeshri, A. Sehaa: A Big Data Analytics Tool for Healthcare Symptoms and Diseases Detection Using Twitter, Apache Spark, and Machine Learning. *Appl. Sci.* **2020**, *10*, 1398. [CrossRef]
5. Yigitcanlar, T.; Mehmood, R.; Corchado, J.M. Green Artificial Intelligence: Towards an Efficient, Sustainable and Equitable Technology for Smart Cities and Futures. *Sustainability* **2021**, *13*, 8952. [CrossRef]
6. Mehmood, R.; Alam, F.; Albogami, N.N.; Katib, I.; Albeshri, A.; Altowaijri, S.M. UTiLearn: A Personalised Ubiquitous Teaching and Learning System for Smart Societies. *IEEE Access* **2017**, *5*, 2615–2635. [CrossRef]
7. Vergunst, F.; Berry, H.L.; Rugkåsa, J.; Burns, T.; Molodynski, A.; Maughan, D.L. Applying the triple bottom line of sustainability to healthcare research—A feasibility study. *Int. J. Qual. Health Care* **2020**, *32*, 48–53. [CrossRef]
8. Eckelman, M.J.; Sherman, J. Environmental Impacts of the U.S. Health Care System and Effects on Public Health. *PLoS ONE* **2016**, *11*, e0157014. [CrossRef]
9. Naylor, C.; Appleby, J. Sustainable Health and Social Care: Connecting Environmental and Financial Performance, p. 28. Available online: https://www.kingsfund.org.uk/publications/sustainable-health-and-social-care (accessed on 12 January 2022).
10. Lenzen, M.; Malik, A.; Li, M.; Fry, J.; Weisz, H.; Pichler, P.-P.; Chaves, L.S.M.; Capon, A.; Pencheon, D. The environmental footprint of health care: A global assessment. *Lancet Planet. Health* **2020**, *4*, e271–e279. [CrossRef]
11. Sherman, E. U.S. Health Care Spending Hit $3.65 Trillion in 2018. *Fortune* **2019**. Available online: https://fortune.com/2019/02/21/us-health-care-costs-2/ (accessed on 12 June 2020).
12. AlOmari, E.; Katib, I.; Albeshri, A.; Mehmood, R. COVID-19: Detecting Government Pandemic Measures and Public Concerns from Twitter Arabic Data Using Distributed Machine Learning. *Int. J. Environ. Res. Public Health* **2021**, *18*, 282. [CrossRef]
13. Ruger, J.P. Health and social justice. *Lancet* **2004**, *364*, 1075–1080. [CrossRef]
14. Jameton, A.; Pierce, J. Environment and health: 8. Sustainable health care and emerging ethical responsibilities. *Can. Med Assoc. J.* **2001**, *164*, 365–369. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC80732/ (accessed on 12 January 2022).
15. Global Social Media Statistics Research Summary. 2022. Available online: https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/ (accessed on 9 January 2022).
16. McColl-Kennedy, J.R.; Vargo, S.L.; Dagger, T.S.; Sweeney, J.C.; van Kasteren, Y. Health Care Customer Value Cocreation Practice Styles. *J. Serv. Res.* **2012**, *15*, 370–389. [CrossRef]
17. Alomari, E.; Katib, I.; Albeshri, A.; Yigitcanlar, T.; Mehmood, R. Iktishaf+: A Big Data Tool with Automatic Labeling for Road Traffic Social Sensing and Event Detection Using Distributed Machine Learning. *Sensors* **2021**, *21*, 2993. [CrossRef] [PubMed]
18. Yigitcanlar, T.; Kankanamge, N.; Regona, M.; Maldonado, A.; Rowan, B.; Ryu, A.; DeSouza, K.C.; Corchado, J.M.; Mehmood, R.; Li, R.Y.M. Artificial Intelligence Technologies and Related Urban Planning and Development Concepts: How Are They Perceived and Utilized in Australia? *J. Open Innov. Technol. Mark. Complex.* **2020**, *6*, 187. [CrossRef]
19. Suma, S.; Mehmood, R.; Albeshri, A. Automatic Detection and Validation of Smart City Events Using HPC and Apache Spark Platforms. In *Smart Infrastructure and Applications: Foundations for Smarter Cities and Societies*; Springer: Cham, Switzerland, 2020; pp. 55–78.
20. Suma, S.; Mehmood, R.; Albugami, N.; Katib, I.; Albeshri, A. Enabling Next Generation Logistics and Planning for Smarter Societies. *Procedia Comput. Sci.* **2017**, *109*, 1122–1127. [CrossRef]

21. Yigitcanlar, T.; Regona, M.; Kankanamge, N.; Mehmood, R.; D'Costa, J.; Lindsay, S.; Nelson, S.; Brhane, A. Detecting Natural Hazard-Related Disaster Impacts with Social Media Analytics: The Case of Australian States and Territories. *Sustainability* **2022**, *14*, 810. [CrossRef]
22. Alsulami, M.; Mehmood, R. Sentiment Analysis Model for Arabic Tweets to Detect Users' Opinions about Government Services in Saudi Arabia: Ministry of Education as a case study. In *Al Yamamah Information and Communication Technology Forum*; Al Yamamah University: Riyadh, Saudi Arabia, 2018; pp. 1–8. Available online: https://www.researchgate.net/publication/324000226_Sentiment_Analysis_Model_for_Arabic_Tweets_to_Detect_Users\T1\textquoteright_Opinions_about_Government_Services_in_Saudi_Arabia_Ministry_of_Education_as_a_case_study (accessed on 7 March 2020).
23. Alomari, E.; Mehmood, R.; Katib, I. Sentiment Analysis of Arabic Tweets for Road Traffic Congestion and Event Detection. In *Smart Infrastructure and Applications: Foundations for Smarter Cities and Societies*; Springer International Publishing: Cham, Switzerland, 2020; pp. 37–54.
24. Alomari, E.; Katib, I.; Mehmood, R. Iktishaf: A Big Data Road-Traffic Event Detection Tool Using Twitter and Spark Machine Learning. *Mob. Netw. Appl.* **2020**, 1–16. [CrossRef]
25. Leite, H.; Hodgkinson, I.R. Telemedicine co-design and value co-creation in public health care. *Aust. J. Public Adm.* **2021**, *80*, 300–323. [CrossRef]
26. Jahanbin, K.; Rahmanian, F.; Rahmanian, V.; Jahromi, A.S. Application of Twitter and web news mining in infectious disease surveillance systems and prospects for public health. *GMS Hyg. Infect. Control* **2019**, *14*, Doc19. [CrossRef]
27. Wang, J.; Wei, L. Fear and Hope, Bitter and Sweet: Emotion Sharing of Cancer Community on Twitter. *Soc. Media Soc.* **2020**, *6*. [CrossRef]
28. Crannell, W.C.; Clark, E.; Jones, C.; James, T.A.; Moore, J. A pattern-matched Twitter analysis of US cancer-patient sentiments. *J. Surg. Res.* **2016**, *206*, 536–542. [CrossRef]
29. Modave, F.; Zhao, Y.; Krieger, J.; He, Z.; Guo, Y.; Huo, J.; Prosperi, M.; Bian, J. Understanding Perceptions and Attitudes in Breast Cancer Discussions on Twitter. *Stud. Health Technol. Inform.* **2019**, *264*, 1293–1297. [CrossRef] [PubMed]
30. Morzy, T.; Härder, T.; Wrembel, R. Advances in Databases and Information Systems. In *Advances in Intelligent Systems and Computing*; Springer: Berlin/Heidelberg, Germay, 2013; Volume 186. [CrossRef]
31. Rasool, M.J.; Brar, A.S.; Kang, H.S. Risk prediction of breast cancer from real time streaming health data using machine learning. *Int. Res. J. Mod. Eng. Technol. Sci.* **2020**, *2*, 409–418. [CrossRef]
32. Diddi, P.; Lundy, L.K. Organizational Twitter Use: Content Analysis of Tweets during Breast Cancer Awareness Month. *J. Health Commun.* **2017**, *22*, 243–253. [CrossRef] [PubMed]
33. Silva, C.V.; Jayasinghe, D.; Janda, M. What Can Twitter Tell Us about Skin Cancer Communication and Prevention on Social Media? *Dermatology* **2020**, *236*, 81–89. [CrossRef]
34. Nguyen, J.; Gilbert, L.; Priede, L.; Heckman, C. The Reach of the "Don't Fry Day" Twitter Campaign: Content Analysis. *JMIR Dermatol.* **2019**, *2*, e14137. [CrossRef]
35. Sedrak, M.S.; Cohen, R.B.; Merchant, R.M.; Schapira, M.M. Cancer Communication in the Social Media Age. *JAMA Oncol.* **2016**, *2*, 822–823. [CrossRef]
36. Sutton, J.; Vos, S.C.; Olson, M.K.; Woods, C.; Cohen, E.L.; Ben Gibson, C.; Phillips, N.E.; Studts, J.L.; Eberth, J.M.; Butts, C.T. Lung Cancer Messages on Twitter: Content Analysis and Evaluation. *J. Am. Coll. Radiol.* **2018**, *15*, 210–217. [CrossRef]
37. Sedrak, M.S.; Salgia, M.M.; Bergerot, C.D.; Ashing-Giwa, K.; Cotta, B.N.; Adashek, J.J.; Dizman, N.; Wong, A.R.; Pal, S.K.; Bergerot, P.G. Examining Public Communication About Kidney Cancer on Twitter. *JCO Clin. Cancer Inform.* **2019**, *3*, 1–6. [CrossRef]
38. Nejad, M.Y.; Delghandi, M.S.; Bali, A.O.; Hosseinzadeh, M. Using Twitter to raise the profile of childhood cancer awareness month. *Netw. Model. Anal. Health Inform. Bioinform.* **2019**, *9*, 3. [CrossRef]
39. Meeking, K. Patients' experiences of radiotherapy: Insights from Twitter. *Radiography* **2020**, *26*, e146–e151. [CrossRef]
40. Zhang, L.; Hall, M.; Bastola, D. Utilizing Twitter data for analysis of chemotherapy. *Int. J. Med. Inform.* **2018**, *120*, 92–100. [CrossRef] [PubMed]
41. Alghamdi, A.; Abumelha, K.; Allarakia, J.; Al-Shehri, A. Conversations and Misconceptions about Chemotherapy in Arabic Tweets: Content Analysis. *J. Med. Internet. Res.* **2020**, *22*, e13979. [CrossRef] [PubMed]
42. Allahyari, M.; Pouriyeh, S.; Kochut, K.; Reza, H. A Knowledge-based Topic Modeling Approach for Automatic Topic Labeling. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 335–349. [CrossRef]
43. Lau, J.H.; Grieser, K.; Newman, D.; Baldwin, T. Automatic labelling of topic models. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Oregon, Portland, 19–24 June 2011; pp. 1536–1545.
44. Wan, X.; Wang, T.; Erk, K.; Smith, N.A. Automatic Labeling of Topic Models Using Text Summaries. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics ACL 2016—Long Papers, Berlin, Germany, 7–12 August 2016; Volume 4, pp. 2297–2305. [CrossRef]
45. He, D.; Wang, M.; Khattak, A.M.; Zhang, L.; Gao, W.; Mateen, A. Automatic Labeling of Topic Models Using Graph-Based Ranking. *IEEE Access* **2019**, *7*, 131593–131608. [CrossRef]
46. Suadaa, L.H.; Purwarianti, A. Combination of Latent Dirichlet Allocation (LDA) and Term Frequency-Inverse Cluster Frequency (TFxICF) in Indonesian text clustering with labeling. In Proceedings of the 2016 4th International Conference on Information and Communication Technology (ICoICT), Bandung, Indonesia, 25–27 May 2016; Volume 4. [CrossRef]

47. Hidayat, T.; Suhardi; Kurniawan, N.B. Smart city service system engineering based on microservices architecture: Case study: Government of tangerang city. In Proceedings of the 2017 International Conference on ICT For Smart Society (ICISS), Tangerang, Indonesia, 18–19 September 2017; Volume 2018, pp. 1–7. [CrossRef]
48. D'Aniello, G.; Gaeta, M.; Orciuoli, F.; Sansonetti, G.; Sorgente, F. Knowledge-Based Smart City Service System. *Electronics* **2020**, *9*, 965. [CrossRef]
49. Yang, Z.; Kankanhalli, A. Innovation in Government Services: The Case of Open Data. *IFIP Adv. Inf. Commun. Technol.* **2013**, *402*, 644–651. [CrossRef]
50. Lopes, A.J.; Pineda, R. Service Systems Engineering Applications. *Procedia Comput. Sci.* **2013**, *16*, 678–687. [CrossRef]
51. Lockheed Martin Corporation. *Guide to the Systems Engineering Body of Knowledge—G2SEBoK*; A Publication of The International Council of Systems Engineering: San Diego, CA, USA, 2004.
52. Demirkan, H.; Spohrer, J.C.; Krishna, V. *Service Systems Implementation*; Springer Science & Business Media: Boston, MA, USA, 2011. [CrossRef]
53. Kalliokoski, P.; Andersson, G.; Salminen, V.; Hemilä, J. BestServ Feasibility Study: Final report. Teknologiainfo Teknovaoy. 2003. Available online: https://cris.vtt.fi/en/publications/bestserv-feasibility-study-final-report. (accessed on 7 March 2020).
54. Musso, M.; Pinna, R.; Melis, G.; Carrus, P. How Social Media Platform can Support Value Cocreation Activities in Healthcare. *Undefined* **2018**, 536–555.
55. Lusch, R.F.; Vargo, S.L. Historical Perspectives on Service-Dominant Logic. *Serv. Log. Mark.* **2021**, 47–60. [CrossRef]
56. Vargo, S.L.; Lusch, R.F. Inversions of service-dominant logic. *Mark. Theory* **2014**, *14*, 239–248. [CrossRef]
57. Lusch, R.F.; Vargo, S.L. The service-dominant logic of marketing: Reactions, reflections, and refinements. *Mark. Theory* **2006**, *6*, 281–288. [CrossRef]
58. Mortenson, M.J.; Vidgen, R. A computational literature review of the technology acceptance model. *Int. J. Inf. Manag.* **2016**, *36*, 1248–1259. [CrossRef]
59. Al Jalim, F. Wasel Electronic Newspaper. Removal of Cancerous Tumor in King Faisal Complex in Taif. Available online: https://wasel-news.com/?p=490731/ (accessed on 20 November 2021).
60. Breastcancer.org. Breast Cancer Facts and Statistics. Available online: https://www.breastcancer.org/facts-statistics/ (accessed on 25 November 2021).
61. Ministry of Health. Breast Cancer Awareness Month. Available online: https://www.moh.gov.sa/HealthAwareness/HealthDay/2018/Pages/HealthDay-2018-10-01-31.aspx/ (accessed on 28 November 2021).
62. Bibri, S.E. Data-driven smart sustainable cities of the future: An evidence synthesis approach to a comprehensive state-of-the-art literature review. *Sustain. Futures* **2021**, *3*, 100047. [CrossRef]
63. Yigitcanlar, T.; Corchado, J.M.; Mehmood, R.; Li, R.Y.M.; Mossberger, K.; Desouza, K. Responsible Urban Innovation with Local Government Artificial Intelligence (AI): A Conceptual Framework and Research Agenda. *J. Open Innov. Technol. Mark. Complex.* **2021**, *7*, 71. [CrossRef]
64. Arfat, Y.; Suma, S.; Mehmood, R.; Albeshri, A. Parallel Shortest Path Big Data Graph Computations of US Road Network Using Apache Spark: Survey, Architecture, and Evaluation. In *Smart Infrastructure and Applications Foundations for Smarter Cities and Societies*; Springer: Cham, Switzerland, 2020; pp. 185–214.
65. Arfat, Y.; Usman, S.; Mehmood, R.; Katib, I. Big data for smart infrastructure design: Opportunities and challenges. In *Smart Infrastructure and Applications Foundations for Smarter Cities and Societies*; Springer: Cham, Switzerland, 2020; pp. 491–518.

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
www.mdpi.com

Sustainability Editorial Office
E-mail: sustainability@mdpi.com
www.mdpi.com/journal/sustainability

MDPI

Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-0093-3