Informatik aktuell

Christian Zinke-Wehlmann
Julia Friedrich *Hrsg.*

First Working Conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow

AI Tomorrow 2023

OPEN ACCESS

Springer Vieweg

Informatik aktuell

Reihe herausgegeben von
Gesellschaft für Informatik e.V. (GI)

Ziel der Reihe ist die möglichst schnelle und weite Verbreitung neuer Forschungs- und Entwicklungsergebnisse, zusammenfassender Übersichtsberichte über den Stand eines Gebietes und von Materialien und Texten zur Weiterbildung. In erster Linie werden Tagungsberichte von Fachtagungen der Gesellschaft für Informatik veröffentlicht, die regelmäßig, oft in Zusammenarbeit mit anderen wissenschaftlichen Gesellschaften, von den Fachausschüssen der Gesellschaft für Informatik veranstaltet werden. Die Auswahl der Vorträge erfolgt im allgemeinen durch international zusammengesetzte Programmkomitees.

Christian Zinke-Wehlmann · Julia Friedrich
Hrsg.

First Working Conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow

AI Tomorrow 2023

Springer Vieweg

Hrsg.
Christian Zinke-Wehlmann
Institut für Angewandte Informatik e.V.
Leipzig University
Leipzig, Sachsen, Deutschland

Julia Friedrich
Institut für Angewandte Informatik e.V.
Leipzig University
Leipzig, Sachsen, Deutschland

ISSN 1431-472X ISSN 2628-8958 (electronic)
Informatik aktuell
ISBN 978-3-658-43704-6 ISBN 978-3-658-43705-3 (eBook)
https://doi.org/10.1007/978-3-658-43705-3

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der DeutschenNationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Springer Vieweg ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature.
Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Das Papier dieses Produkts ist recycelbar.

Inhaltsverzeichnis

AI supported knowledge management and learning

Manufacturing and factories of the future

Introduction

Christian Zinke-Wehlmann(✉) and Julia Friedrich

Institut für Angewandte Informatik (InfAI) e. V., Kompetenzzentrum Künstlich Menschlich Intelligenz, Leipzig, Deutschland

{zinke,friedrich}@infai.org

Zusammenfassung. Technologische Innovationen verändern die Art und Weise, wie Gesellschaften funktionieren. Sie bieten bspw. einen besseren Zugang zu Wissen, eine individuellere Produktentwicklung, ermöglichen es, neue Geschäftsfelder zu erschließen oder Bürokratie abzubauen. Folglich geht die digitale Transformation in Organisationen über die bloße Übernahme innovativer Technologien wie künstliche Intelligenz hinaus. Sie erstreckt sich auf die gesamte Organisation, einschließlich ihrer Strukturen, Prozesse und Mitarbeitender. Besonders kleine und mittlere Unternehmen (KMU) stehen von der Herausforderung diese Technologien zu adaptieren und Möglichkeiten zu finden, diese mit zu gestalten - oft fehlen hier aber die Ressourcen. Jüngste Entwicklungen wie die EU-Verordnungen über künstliche Intelligenz (KI) und Krisen wie unterbrochene Lieferketten, Fachkräftemangel und Klimawandel haben die Situation für Unternehmen noch schwieriger gemacht. Das Konzept der Industrie 5.0, das als nächster evolutionärer Schritt in der Entwicklung der Industrie angesehen wird, zielt darauf ab, diese Anforderungen zu erfüllen. Industrie 5.0 wird als zukunftssicherer, widerstandsfähiger, nachhaltiger und menschenzentrierter beschrieben. Menschenzentrierte Ansätze sollten als Haupttreiber und Innovationsfaktor für die Industrie 5.0 anerkannt werden. Ein menschenzentrierter Ansatz für die digitale Transformation erfordert, die Bedürfnisse und Interessen der Menschen stärker zu berücksichtigen anstatt alleinig auf die Technologie zu fokussieren. Das Ziel der AI Tomorrow Konferenz ist es, diesen Diskurs zu fördern, indem die nachhaltige und menschenzentrierte Perspektive der digitalen Transformation in den Mittelpunkt der wissenschaftlichen Diskussion gestellt werden. Die erste AI Tomorrow-Konferenz bot Wissenschaftler:innen unterschiedlicher Forschungsdisziplinen Raum, Herausforderungen und Ideen für die Zukunft der KI zu diskutieren.

Robots and algorithms, data and metaverse – these seem to be what the future is made of. While programmers and data scientists build the fundament of a data dominated world, the discussion of ethical and ecological impacts for society and economy often only come in second. We, the editors of this proceeding, think that both, innovation and impact should both be well considered in parallel. Therefore, the scientific community should always keep a holistic view on AI which includes trying to anticipate possible outcomes and consider the impact of AI for the way we live, we work and communicate. With our very first AI Tomorrow conference, we wanted to create a space, where scientists from different fields can come together and discuss their ideas for the future of AI.

C. Zinke-Wehlmann und J. Friedrich (Hrsg.): AIDRST 2023, INFORMAT, S. 1–4, 2024. https://doi.org/10.1007/978-3-658-43705-3_1

The first working conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow (www.ai-tomorrow.org/) took place from 29th to 30th of June 2023. Its aim was to bring together scientists from various fields that focus on more than the technological aspects of AI development. We wanted to shed light on the opportunities and challenges of AI with regard to resilience and sustainability of business models, societies and the environment.

Digital innovations change the way in which societies work. On the one hand, the increasing amount of digital devices and their performance capacity bring new opportunities to all aspects of our live, e.g. in terms of travel planning due to smart route planning and ticket booking, the reduction of red tape, e.g. through digital identity cards, and or even inclusion processes, e.g. through text-to-speech applications or semi-automated translation or summarizing of texts. Likewise, the increasing number of AI-based tools and methodologies change the world of businesses. This includes improved knowledge access, more extensive product development and the creation of new business areas. The speed in which processes and technologies change is challenging for all. Companies all over the world must adapt to the rapid development, which requires a multidisciplinary approach to digital transformation to remain competitive.

One particular focus of our conference is on small and medium-sizes companies (SME). In Germany, SME have a major impact on the economy. In fact, the share of SMEs in Germany was 99.4 percent in 2019.[1] They account for the majority of all apprenticeships, making them one of the main drivers of competence development.[2] While large companies have the financial abilities to hire change manager and invest in consulting and training, SMEs struggle to keep up with the speed of the technological development because they simply don't have the financial and personal resources for adaption. To compete with larger businesses in terms of affordable prices and high-quality products, SMEs need to adapt to new digital innovations. The use of artificial intelligence (AI) – one of the key technologies of digitalization – can have a significant impact on business success and is increasingly becoming a game changer for SMEs. In contrast, Accenture's 2022 study[3] shows that only twelve percent of companies consider their AI maturity level as high enough to use it to grow and transform their business, while 63 percent are still in the experimental phase of adopting AI in the enterprise.

1 Statistisches Bundesamt (2019): Shares of small and medium-sized enterprises in selected variables. URL: https://www.destatis.de/EN/Themes/Economic-Sectors-Enterprises/Enterprises/Small-Sized-Enterprises-Medium-Sized-Enterprises/Tables/total-cik.html, last accessed 1st September 2023.

2 Federal Ministry for Economic Affairs and Energy (2019): SMEs digital. Strategies for the digital transformation. URL: https://www.bmwk.de/Redaktion/EN/Publikationen/Mittelstand/smes-digital-strategies-for-digital-transformation.pdf?__blob=publicationFile&v=5, last accessed 1st September 2023.

3 Accenture (2022): The art of AI maturity: Advancing from practice to performance. URL: https://www.accenture.com/_acnmedia/Thought-Leadership-Assets/PDF-5/Accenture-Art-of-AIMaturity-Report.pdf, last accessed 1st September 2023.

In addition to the challenges of digital transformation, current developments and crises such as EU AI regulations, disrupted supply chains, skilled workers shortage and climate change, have broadened the focus and question the effects and opportunities of digitization with a view to the next evolutionary step in the development of Industry 4.0, namely Industry 5.0. In a 2021 policy brief, the European Commission describes Industry 5.0 as “more future-proof, resilient, sustainable, and people-centric”.[4] While Industry 4.0 increases resilience and “the positive impact on the ecological and economic dimensions seems to be fairly widespread”, the social dimension, especially human-centric approaches and measures, may still be underestimated. People or human-centric approaches are therefore the main driver and innovation factor for Industry 5.0. A human-centered approach to digital transformation means focusing on people’s needs and interests rather than viewing technology as the sole actor.

With this focus on human-centred approaches in AI development; a multi-dimensional value discussion about impacts and sustainability of AI; measurements for AI values; concrete use cases of AI development (in SME) the AI Tomorrow Conference brings experts from various fields together. It is the goal of the conference and our hope to foster an interdisciplinary dialogue and identify opportunities for collaboration and advancement in the responsible development and deployment of AI.

The findings of the articles in this proceeding have been presented at the 1st Conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow (AI Tomorrow) which was hold an 29th and 30th of June 2023 in Leipzig. A special thanks go to the session chairs, Dr. Sandra Schumann, Anja Brückner, Dr. Amit Kirschenbaum and Jasmin Mayan. Their support in form of moderation and organization of their sessions was a crucial factor in making the 1st AI Tomorrow Conference a success.

Except for the positioning paper which sums up the first conference keynote, given by Dr. Christian Zinke-Wehlmann, the articles in the proceedings have been peer reviewed by the members of the scientific board in a double-blind review process.

Leipzig, Germany
28.08.2023

[4] European Commission, Directorate General for Research and Innovation (2021): Industry 5.0: towards a sustainable, human centric and resilient European industry. Publications Office of the European Union.

Resilient and Sustainable AI. Positioning paper on the relation of AI, resilience and sustainability

Dr.Christian Zinke Wehlmann(✉)

Institut für Angewandte Informatik (InfAI) e. V., Kompetenzzentrum Künstlich Menschlich Intelligent, Goerdelerring 9, 04109 Leipzig, Germany
zinke@infai.org

Abstract. In the contemporary debate, surrounding the future of work and life, Artificial Intelligence (AI), resilience, and sustainability have emerged as pivotal concepts. Within the industrial realm, their collective convergence is driving unprecedented transformative shifts, challenging traditional paradigms. This positioning paper delves into the intricate interlinkages binding these three paradigms. Examples such as AI-driven automation, enhancing efficiency, and predictive maintenance, reducing machinery downtime, underscore the transformative role of AI in the industry. Meanwhile, an increasing emphasis on environmental responsibility highlights the growing importance of sustainability in the industrial sector. Resilience, embodied through the ability to withstand crises and maintain strong supply chains, is equally essential. The article also delves deep into the specific relations between AI, sustainability and resilience. By weaving these concepts together, the paper aims to provide a holistic perspective on the interconnectedness, emphasizing the need for a balanced approach in the modern industry to ensure not only technological advancement but also a resilient and sustainable future.

Keywords: Artificial Intelligence · Resilience · Sustainability

1 Introduction

Artificial intelligence, resilience, and sustainability are major concepts of the current debate on the future of work and life. In the industrial landscape, the convergence of AI, sustainability, and resilience is ushering in transformative changes that will reshape traditional paradigms. Each of these elements itself, and to an even greater extent together, are emerging as key drivers for the next era of industrial growth. Some well-known examples of the **pivotal role of AI in Industry are:**

- Automation: AI-driven automation streamlines processes, reduces waste, and optimizes resource use, and thus leading to enhanced efficiency and productivity. E.g. a study of Brynojolfsson et al. [8] found that generative AI increases the work of customer support agents significantly.

C. Zinke-Wehlmann und J. Friedrich (Hrsg.): *First Working Conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow*, AIDRST 2023, Informatik aktuell, S. 5–19, 2024. https://doi.org/10.1007/978-3-658-43705-3_2

- Predictive maintenance: Through analysis, AI identifies and predicts potential machinery breakdowns before they occur, reducing downtime and ensuring smooth operations [11].

An example for the crucial impact of sustainability on industry is:

- Environmental Responsibility: Industrial sectors are under increasing pressure to align with global environmental goals. Implementing sustainable practices can drastically reduce the carbon footprint and resource wastage. A report by the Business and Sustainable Development Commission [9] highlighted that embracing sustainability could unlock $12 trillion in new market value by 2030.

To underline the concept of Resilience, well know examples of usage are:

- Withstand Crisis: In an ever-changing global landscape filled with economic uncertainties, geopolitical tensions, and the looming threat of pandemics, industries must prioritize resilience to manage risks. According to Deloitte [12], resilient organizations[1] are 3 times more likely to outperform their industry peers during a crisis.
- Supply Network Strength: Resilience ensures continuity in supply chains, even amidst disruptions. A resilient system can quickly adapt and recover from shocks, exemplarily emphasized by [31].

For the modern industry, AI, sustainability, and resilience aren't just buzzwords – they are essential pillars that will define their growth, reputation, and longevity in the coming decades. As these concepts may be deeply interwoven, industries must prioritize their simultaneous development and integration. This article introduces these three concepts and analyses the way in which they are linked together. Examples for this interlinkage are AI algorithms that supports sustainable energy industry [1] or AI-driven data analytics that can be pivotal in developing sustainable supply chains by predicting potential disruptions and ensuring a stable, resilient system [22]. It is no aim to provide a comprehensive overview of the state of research, but to sharpen the terms and highlight their mutual interdependencies.

Therefore, this positioning paper delves into the intricate relationships binding AI, resilience, and sustainability. We aim to elucidate the underpinnings of resilient and sustainable AI systems – with a holist point of view [26] – and finally provide a cohesive point of view that interlinks all three paradigms. By charting this tripartite connection, we hope to lay a foundation for future research and guide AI's trajectory towards a future that is not just technologically advanced, but also resilient and sustainable.

[1] The 2021 Deloitte Global Resilience Report summarizes attributes like preparedness, adaptability, collaboration, trustworthiness, and responsibility for resilient organizations.

Before we can describe the interlinkage between the main concepts, we need to define the terms. After introducing the term AI, we will have a closer look on resilience as well sustainability.

2 What is AI?

Even though everybody talks or writes about AI, most people have no concrete idea of what the term actually stands for. To talk about artificial intelligence, it is therefore necessary to define it in the first place. In the current period, with new software products based on machine learning (ML) technology appearing almost daily, ML dominates the public's perception and broad understanding of what AI is.

In a JRX technical report by the EU [16], various definitions were taken up and analysed by AI Watch with the goal of sharpening the vague concept of AI. It concludes in stating, that AI is a catch-all term and can mean different things:

"AI is a **generic term** that refers to any machine or algorithm that is capable of

- observing its environment,
- learning,
- and based on the knowledge and experience gained, taking intelligent action or proposing decisions." [16]

This definition approach essentially builds on the idea of a description of the capabilities of the technology. This means that the AI term can thus encompass any technology that is capable of observing, learning, and, based on built-up knowledge, making decisions, and recommending or implementing them. Hence, we can see that multiple technologies fall under such a broad AI definition.

AI Watch develops an AI taxonomy based on this definition to cluster various AI approaches. Core areas of AI (Reasoning, Planning, Learning, Communication, Perception) are distinguished from transversal areas (Integration & Interaction, Services, Ethics & Philosophy) and sub-areas are named, such as knowledge representation in the area of reasoning [16]. Beside the more or less academic discourse about the question "What is AI?", AI is getting to practice, which results in a political as well as legal definition of AI.

In the course of developing rules for AI ("LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS") by the EU, three technologies are fundamentally distinguished:

"(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;

(b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
(c) Statistical approaches, Bayesian estimation, search and optimization methods." [15]

Both presented definitions are highly overlapping and demonstrate somehow the close relation between legal/political interpretation and a specific academic debate. Again, this paper does not aim to present a full history of AI or a literature-based analysis of the current AI definition debate. From the authors point of view, the presented definitions are a practical ground for current developments of AI (Europe) in research and industry because of its role as legal framework. It sets the basics to understand that the term "AI" is generic, but refers to specific properties of technical systems, which can manifest completely differently – as a robot or as a statistical model. The expressions and intensity of the abilities are not further defined here. It describes what a technical system does (observe, learn, propose or implement decisions based on collected knowledge) to determine if this system is artificially intelligent.

3 What is Resilience?

Resilience is a term that is used in many contexts, e. g. psychology. It probably originally stems from the field of physics [25]. In the following, we will refer to economic resilience, which, likewise, can have many different forms. Resilience occurs within companies, as well as within value-added networks, and entire economies. The term was first used by Holling; stating that economic resilience "refers to the extent to which a shock can be absorbed by a local stable domain before it is induced into some other stable equilibrium" (Modica & Reggiani, 2015). Looking at the term through a systemic lens (e.g., in addition to 6), resilience is the ability of a system,

- to ensure the system's stability despite (massive) external (or internal) disturbances/crises,
- to adapt to permanent changes caused by crises.

Similar to the AI definitions, the given approach to define resilience bases on an ability of a system (in this case, not necessarily a technical system). Unlike attributing the *what* a system must do (see AI), this definition rather focuses on the consequences or the result. This definition stands in contrast to the definition of organizational resilience in standardization. E.g., in the ISO standards' definition, resilience isn't an element of corporate control but emerges from the application of these control elements. [19] The question is whether a result of applying unspecified mechanisms can be a capability of systems? The aforementioned definition suggests precisely this – but above all, it suggests that systems can exhibit mechanisms that produce resilience. To address this challenge, resilience can be defined as responsiveness:

"Resilience is the responsiveness of the system, i.e., its elasticity or capacity to rebound after a shock, indicated by the degree of flexibility, persistence of key functions, or ability to transform […]" [27]

The mechanisms to achieve this can be termed as a resilience strategy. This, essentially, describes the deliberate approach to achieving resilience and can be clustered into three areas:

- Preparation and prevention of shocks/disturbances or crises,
- Mitigation and prevention of consequences from the disturbance/crisis,
- Adjustments to the crisis (consequences of the crisis).

If a system possesses mechanisms that serve this purpose and are functional, then a system can be described as resilient. Such mechanisms are, for instance, part of supply chain research, which, for example, views "Situation Awareness" as a metric for "Readiness" as preparation see [19]. On the other hand, there are a number of standards that deal with resilience strategies. For companies, there's ISO 22316, which suggests what to do or exactly how to achieve resilience, e.g., that the management should establish relevant business resilience goals and formulate assessment criteria for evaluating resilience attributes [21]

In addition to these standards, there are several others, closely tied to this capability, such as Compliance Management System (CMS) according to ISO 37301, Risk Management: ISO 31000, Knowledge Management: DIN EN ISO/IEC 27.001:2017; ISO 30301:2019; DIN ISO 30401:2021, Supply Chains: ISO 28000; DIN ISO 20400:2021 ISO/TS 22318. As already explained in the introductory words, this presentation is not comprehensive but provides a brief insight without even approaching the state of research.

4 What is Sustainability?

After AI and resilience, sustainability remains to be addressed, an equally complex term. From the authors perspective, the concept of sustainability only becomes meaningful when it is understood as a sustainable development, meaning a process – and this is also how some other authors perceive it. Sustainability itself is a "paradigm for thinking about the future in which environmental, societal, and economic considerations are balanced in the pursuit of improved quality of life. The ideals and principles behind it rest on broad concepts such as intergenerational equity, gender equity, social tolerance, poverty alleviation, environmental preservation and restoration, natural resource conservation, and building just and peaceful societies" [23].

In the context of the so-called Triple Bottom Line and from a business perspective, sustainability is often understood as follows:

- Sustainability seeks to evaluate the corporation's financial, social, and environmental outcomes over a duration. Only an enterprise that adopts a TBL approach truly acknowledges the comprehensive expenses associated with its operations. [13]

However, the "inventor" of the TBL, John Elkington, clarified in 2018 that, "[…] the original idea was broader still, urging businesses to track and manage economic (not just financial), social, and environmental value added – or destroyed." [14] Thus, sustainability can't just be measured by profit and loss, but by the well-being of people, now and in the future. The TBL emphasized radical transformation, market disruption, asymmetric growth (wherein non-sustainable sectors are deliberately marginalized), and the amplification of emergent market strategies (Elkington, 2018). Moreover, the TBL can be understood as a hierarchical system, not an equal one, because: Ecological conditions and resources serve as foundational elements for both life and societal structures. A well-operating society is pivotal for addressing social necessities and instituting an economic framework. This economic structure facilitates both resource utilization and prosperity, catering to material requirements [20].

From the political debate on sustainability, the concept of sustainable development marks a milestone. Sustainable development is a process filled with life in the context of the UN, especially the Brundtland Commission, which states that sustainable development addresses contemporary requirements while ensuring that future generations retain the capacity to fulfill their respective needs [5]. Without delivering a complete historical account (there's plenty of literature for that, e.g., [18]), these developments culminated in the Sustainable Development Goals, the UN's 17 sustainability targets. These are:

#1 No Poverty
#2 No Hunger
#3 Good Health and Well-being
#4 Quality Education
#5 Gender Equality
#6 Clean Water and Sanitation
#7 Affordable and Clean Energy
#8 Decent Work and Economic Growth
#9 Industry, Innovation, and Infrastructure
#10 Reduced Inequalities
#11 Sustainable Cities and Communities
#12 Responsible Consumption and Production
#13 Climate Action
#14 Life Below Water
#15 Life on Land
#16 Peace, Justice, and Strong Institutions
#17 Partnerships for the Goals

Combining these objectives with the TBL, goals 6, 13, 14, 15 (environment) can be seen as the foundation for goals 1, 2, 3, 4, 5, 7, 11 and 16 (society), which in turn enable goals 8, 9, 10, 12 (economy) and can be achieved through 17 (cooperation). As an example gender equality (#5) can be seen as a driver for high quality education (#4), enabling decent work and economic (#8) in the long term [24]. However, a too simplistic understanding would neglect the high interdependence of individual factors, as cause and effect relations are often precisely the opposite, e.g., sustainable production allows life on land and so on.

Having introduced the third term, the next sections will relate two terms to each other, finally focusing on the relationship between AI, sustainability, and resilience and discussing the role of AI.

5 Sustainability and Resilience

Resilience and sustainability seem somewhat contradictory in the presented version. While sustainable development takes a very long-term view, resilience is more focused on narrow time horizons. This is especially the case when we correctly consider the type of response. Resilience is a reactive ability to a very abrupt disruption or crisis, unlike sustainable development which focuses on changes observed over very long processes. However, both concepts consider the maintenance of systems and their performance, but the time horizons they consider are very different.

Nevertheless, both concepts are related. A sustainable system should be able to handle shocks and crises. “In political interpretation, resilience and sustainability are two sides of the same coin. Only sustainable value chains remain robust and intact in the long run. It is therefore essential that companies keep an eye on the impact on people, the environment, and the climate in their activities” (4, translated). The logical relationship that arises from this is: “Resilience represents a necessary, but not sufficient, condition for sustainability.” (7, translated). The logical relationship that arises from this is: “Resilience represents a necessary, but not sufficient, condition for sustainability” (Translated from 7).

In short: Not every resilient system is sustainable, but every sustainable system is resilient. This dependency and interdependence can lead to elements of sustainable development, being part of resilience strategies as they can be found in the BMWK-Whitepaper in context of Platform Industry 4.0 [3]. Sustainability is one of the levels of the resilience strategy, which spans the various impact phases of resilience. The social dimension of sustainability – health protection, societal benefit, working conditions – is particularly emphasized. In other words, in Industry 4.0, social sustainability constitutes a sufficient condition for resilience

6 AI and Resilience

Resilience and AI can be related in various ways. As already described, AI always pertains to a technical system’s ability to learn/act, while resilience addresses a systems responsiveness. From this, three potential connections arise:

- AI systems are used to establish this resilience, making AI an integral part of the resilience strategy.
- Developments in AI systems can cause a shock to companies or industries.
- The resilience of AI systems itself comes into focus.

Regarding the point that AI is used for resilience, there are numerous (German or EU) research projects and publications (e.g. [2, 10, 31]). Projects and their areas can be taken from the following table (Tab. 1):

Table 1. Resilience and AI – examples of domains and research projects

Infrastructure	aKtIv: Agile network control to increase the resilience of the critical infrastructure water supply BESKID: Fire design simulation in rail vehicles using AI-based data IKIGas: Industrial Artificial Intelligence for safety in gas networks
Society	NEBULA: User-centered AI-based detection of fake news and misinformation AIFER: Artificial intelligence for analyzing and merging Earth observation and internet data for decision support in disaster management FAKE-ID: Video analysis using artificial intelligence to detect false and manipulated identities
Supply chains and organizations	LEAS: Land-side recommendation for traffic situations with highly automated or autonomous ships SPAICER: Minimize production disruptions and interruptions in supply chains ResKriVer – Communication & Information platform for resilient, crisis-relevant supply networks PAIRS – Privacy-Aware, Intelligent and Resilient Crisis Management KISS – AI-supported Rapid Supply Network DAKI-FWS – Data- and AI-supported early warning system CoyPu – Cognitive Economy Intelligence Platform for the resilience of economic ecosystems

AI can be used in all phases of the resilience strategy. The most intuitive applications involve AI-supported prediction systems for specific scenarios. This is particularly useful when risks for a shock can be foresight – hence, enough data with enough information is available to make such predictions.

Besides AI's beneficial power to strengthen resilience, AI system may also have other roles. The second role of AI is, that AI is realized as a trigger for shocks on industries or companies. A publicly discussed example of such kind of shock, was the discussion about ChatGPT and its impact on stocks in the education market, which (temporarily) plunged[2]. This example is a prototype for public discourses about the impact of AI, which moves between from euphoria to dystopia (in this case near economic dystopia). Reality probably lies somewhere in the middle, as AI development influences society in general – just as society and economic systems influence techno-

[2] See press: https://www.welt.de/wirtschaft/article245123004/KI-50-Prozent-Absturz-die-erste-Branche-droht-ChatGPT-zum-Opfer-zu-fallen.html#cs-lazy-picture-placeholder-01c4eedaca.png. To my knowledge, there is no scientific analysis of the direct connection or evidence between the two events – however, there is a certain plausibility.

logical developments, see also [33]. The interdependencies are complex, making it not unlikely that AI developments will cause shocks to hit companies and sectors.

Hence, it seems logical to examine how resilience can be built up for such case. Initially, understanding what AI is and that it will induce change should help organizations prepare for potential AI-driven shocks. Building this competence or accessing educational resources (also pertaining to social sustainability), is useful throughout all resilience phases. The level of competency being built also aids during the adaptation phase, leveraging the shock for potential adaptations and restructuring measures.

Lastly, the resilience of AI systems themselves can be examined. Why? Since the beginning of computing, various significant errors have happened due to numerical inaccuracies. A particularly infamous incident was the self-destruction of an Ariane 5 rocket in 1996. Such events have driven the establishment of a dedicated research area that concentrates on creating and refining resilient algorithms, a niche within the realm of numerics. This needs to be extended on AI Systems to prevent the given risks [30].

As AI systems increasingly permeate our daily lives and work, it becomes essential to scrutinize their resilience to prevent potential harm to social systems or the organizations we work in. Wittenbrink et al. [30] tackled this challenge, concluding that five factors define resilient AI:

- Safety
- Accuracy
- Reliability
- Robustness
- Comprehensibility

These points refer, among other things, to the data foundation upon which any AI is based, its origin, and conformity (Governance Risk Compliance or GDPR), IT security, and the model's explainability and resilience, or that of the technical system. Research around this topic is, to the authors knowledge, still in its infancy, and the question of extending and specifying these points is probably an ongoing process. In the future, questions about model transparency and the focus on the common good will likely arise. I will revisit the question of AI system resilience when introducing the connection between AI and sustainability, as resilience and sustainability are logically connected, and some readers might have noticed certain overlaps.

7 Sustainability and AI

There are several works about sustainability and AI. As an example, the definition of sustainable AI is still being formulated. Bjørlo et al. and Halsband focuses on the idea that sustainable AI should meet current needs without compromising the future – in addition to the presented definition of sustainability. Some researchers have identified boundary conditions for sustainable AI, including diversity, trust, and the capacity for self-organization and learning.

One interesting work is "Sustainability Criteria for Artificial Intelligence" („Nachhaltigkeitskriterien für künstliche Intelligenz") authored by Friederike Rohde, Josephin Wagner, Philipp Reinhard, Ulrich Petschow, Andreas Meyer, Marcus Voß, and Anne Mollen [26]. The paper delves into the realm of artificial intelligence (AI), particularly machine learning (ML) and sustainability. The increasing deployment of AI systems has sparked debates globally about their societal, environmental, and economic impacts. Concerns include non-transparent decision-making processes, discrimination, rising energy consumption, greenhouse gas emissions during AI model development, and broader consequences on labor markets, consumption patterns, and the market power of large corporations. The authors aim to provide a perspective on sustainable AI, connecting it with discussions on AI ethics, Green AI, and the sustainability of AI and advocate for an embedded perspective on sustainable AI, viewing technology as socially shaped and shapable, not just a neutral tool [26].

From a sustainability perspective, AI systems present multiple challenges. One of these significant challenges is the interdependency between different SDGs and AI. Direct and indirect impacts of AI on SDGs need to be considered[3]. Vinuesa et al. [28] note that AI could aid in achieving 79% of the Sustainable Development Goals, like alleviating poverty and improving education. However, it could also hinder 35% of these goals by, for instance, consuming vast natural resources, propagating biases against gender equality, or bolstering autocratic regimes. While AI in sectors like smart manufacturing can boost productivity and conserve resources, it might lead to job losses in areas like finance[4]. However, the work of Rohde et al. (2021) includes a systematic overview of impacts throughout the AI lifecycle. Further Rohde et al. presents conceptual ideas for a comprehensive sustainability assessment of AI. It introduces thirteen sustainability criteria and five cross-sectional criteria with corresponding indicators based on existing scientific and societal discourses on the impacts of AI [26]:

Sustainability Criteria for Artificial Intelligence:

Ecological Criteria:

- Energy Consumption: Refers to the amount of energy utilized by AI systems during their operation, and further emphasizing the need for energy-efficient solutions.
- CO2 and Greenhouse Gas Emissions: Highlights the environmental impact of AI systems in terms of carbon dioxide and other greenhouse gas emissions, advocating for low-emission technologies and solutions.
- Sustainability Potentials: Discusses the potential of AI to contribute positively to environmental sustainability, such as optimizing resource use or aiding in conservation efforts.
- Indirect Resource Consumption: Addresses the resources consumed indirectly due to AI operations, such as the materials used in hardware manufacturing or the water footprint of data centers.

[3] First benchmarking initiatives like the platform for learning systems (translated) (https://www.plattform-lernende-systeme.de/nachhaltigkeit-karte.html) underlines the importance.

[4] "Goldman Sachs employed six hundred traders in 2000, the corporation was able to reduce their number of human traders to two by 2017 because of advances in narrow AI" [17].

Social Criteria:

- Transparency & Responsibility/Accuracy: Stresses the need for clarity in how AI systems operate and make decisions, ensuring that users and stakeholders understand and can trust the technology.
- Technical Reliability & Human Oversight: Emphasizes the importance of AI systems being reliable and accurate, with mechanisms for human intervention and oversight when needed.
- Data Protection & Privacy: Emphasizes the protection of user data and privacy, ensuring that AI systems handle personal information responsibly and securely.
- Inclusive and Participatory Design: Advocates for the design and development of AI systems that are inclusive of diverse user groups and allow for stakeholder participation in decision-making processes.
- Cultural Sensitivity: Stresses the importance of AI systems being respectful and considerate of cultural differences, ensuring that they do not perpetuate biases or stereotypes.

Economic Criteria:

- Distribution Effect in Target Markets: Discusses how AI impacts the distribution of resources, wealth, or opportunities in its target markets, emphasizing equitable distribution.
- Working Conditions & Working Place: Highlights the impact of AI on the workplace, including potential changes in job roles, working conditions, and the nature of work.
- Market Diversity & Exploitation of Innovation Potential: Emphasizes the importance of a diverse market landscape where AI fosters innovation without leading to monopolies or stifling competition.

Cross-sectional Criteria:

- Defined Responsibilities: Organizations should have clear responsibilities for ensuring the sustainability of AI.
- Code of Conduct: Defines the values and norms for the implementation and use of AI systems.
- Stakeholder Analysis & Participation: Involves identifying and integrating stakeholders in the AI governance process.
- Documentation of AI Systems: Comprehensive documentation of AI systems, including data sources and methodologies.
- Risk Management: Identifying and managing potential risks associated with AI systems.

With the definition and the naming of criteria for the sustainable use of AI, a first specification has been made that helps all organizations to orient themselves. In the following chapter, we will draw a conclusion and bring together the relationships of the triad AI, resilience, and sustainability.

8 Conclusion: On the relation of AI, resilience and sustainability

In the previous chapters, numoerous definitions and explanation approaches for the understanding of the major concepts of AI, sustainability and resilience and their interconnections were given. To sum it up, we can state that AI is a generic term for technical systems with special capabilities (observe, learn, act), Resilience refers to a responsiveness of a system, and sustainability or sustainable development addresses contemporary requirements while ensuring that future generations retain the capacity to fulfill their respective needs.

AI, understood as a technical system, can be resilient, and in an extended sense, sustainable (with sustainability implying resilience, as demonstrated). Hence, AI systems need to ensure to fulfil or not to violate contemporary requirements on sustainability (given above). Therefore, resilient AI systems are a necessary but not sufficient condition for sustainable AI systems. Technical systems are integral components of socio-technical systems, suggesting that these technical systems impact and are interdependent with social, ecological, and economic systems throughout the AI systems' lifecycle. The sustainability index approach introduced for AI [26] is a pivotal step in identifying and measuring these interdependencies. To integrate the aforementioned works, the author proposes that the sustainability and inherent resilience of AI systems, as well as their impact concerning sustainability factors, are sufficient criteria for evaluating AI's overall sustainability. From this perspective, it's unsurprising that many resilient factors are already evident within the presented sustainable AI index, such as the data foundation in AI systems documentation and Technical Reliability. This suggests that some resilience factors might need more in-depth consideration in terms of [30]:

- Safety (Hardware, Model, Data)
- Robustness (e.g., adversarial robustness)

Moreover, we must recognize that certain AI technologies could be leap innovations (market shocks) and might be unpredictable. We could also face challenges when trying to identify rebound effects. AI technologies will invariably introduce potential shocks to systems. To prepare for these shocks, developing competencies has been showcased as an effective way to navigate such potential AI-induced disruptions (and perhaps even broader ones), thus fostering sustainable socio-technical systems. As a result, the index should also give significantly more weight to: Competency development for co-creating AI systems as a pertinent factor.

Besides those factors, more generally, we may need to considerate and classify the factors in different level. Finally, sustainable AI systems needs to be resilient – it is a logical consequence.

References

1. Ahmad, T., Zhang, D., Huang, C., Zhang, H., Dai, N., Song, Y., & Chen, H. (2021). Artificial intelligence in sustainable energy industry: Status Quo, challenges and opportunities. *Journal of Cleaner Production, 289,* 125834. https://doi.org/10.1016/j.jclepro.2021.125834.

2. Alshater, M. M., Kampouris, I., Marashdeh, H., Atayah, O. F., & Banna, H. (2022). Early warning system to predict energy prices: the role of artificial intelligence and machine learning. *Annals of Operations Research,* 1–37. https://doi.org/10.1007/s10479-022-04908-9.
3. BMWK. (2022). Plattform Industrie 4.0: Resilienz im Kontext von Industrie 4.0.
4. BMWK. (2023). „Resilienz und Nachhaltigkeit gehen Hand in Hand" – OECD-Ministertreffen zu verantwortungsvollem unternehmerischem Handeln. https://www.bmwk.de/Redaktion/DE/Pressemitteilungen/2023/02/20230215-resilienz-und-nachhaltigkeit-gehen-hand-in-hand.html. Zugegriffen: 1. Sept. 2023.
5. Borowy, I. (2013). *Defining sustainable development for our common future: A history of the World Commission on Environment and Development (Brundtland Commission)*. Routledge.
6. Brende, B., & Sternfels, B. (2022). Resilience for sustainable, inclusive growth. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/resilience-for-sustainable-inclusive-growth. Zugegriffen: 1. Sept. 2023.
7. Brinkmann, H., Harendt, C., Heinemann, F., & Nover, J. (2017). Ökonomische Resilienz—Schlüsselbegriff für ein neues wirtschaftspolitisches Leitbild? *Wirtschaftsdienst, 97*(9), 644–650
8. Brynjolfsson, E., Li, D., & Raymond, L. (2023). *Generative AI at work*. National Bureau of Economic Research.
9. Business, Sustainable Development Commission, others. (2017). Better business, better world. Systemiq. https://sustainabledevelopment.un.org/content/documents/2399BetterBusinessBetterWorld.pdf.
10. Charles, V., Emrouznejad, A., & Gherman, T. (2023). A critical analysis of the integration of blockchain and artificial intelligence for supply chain. *Annals of Operations Research,* 1–41. https://doi.org/10.1007/s10479-023-05169-w.
11. Chui, M., Henke, N., & Miremadi, M. (2018). Most of AI's business uses will be in two areas. *Harvard Business Review, 20*.
12. Deloitte. (2021). Building the Resilient Organization. 2021 Deloitte Global Resilience Report. *Deloitte*.
13. Elkington, J. (2009). Triple bottom line| It consists of three Ps: Profit, people and planet. *The Economist*.
14. Elkington, J. (2018). 25 years ago I coined the phrase "triple bottom line." Here's why it's time to rethink it. *Harvard Business Review, 25,* 2–5.
15. European Commission. (2021). Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. *European Commission, 106,* 1–108.
16. European Commission, Joint Research Centre, Samoili, S., López Cobo, M., Delipetrev, B., Martínez-Plumed, F., Gómez, E., & De Prato, G. (2021). AI watch. Defining artificial intelligence 2.0. Towards an operational definition and taxonomy for the AI landscape. *Publications Office of the European Union*.
17. Goralski, M. A., & Tan, T. K. (2020). Artificial intelligence and sustainable development. *The International Journal of Management Education, 18*(1), 100330. https://doi.org/10.1016/j.ijme.2019.100330.
18. Hauff, M. (2021). *Nachhaltige Entwicklung*. De Gruyter.
19. Herdmann, F. (2023). Resilienz in Unternehmen. Beuth DIN. https://www.beuth.de/de/themenseiten/resilienz-in-unternehmen. Zugegriffen: 1. Sept. 2023.
20. Holzbaur, U. (2020). Nachhaltige Entwicklung – Sustainable Development. In U., Holzbaur (Hrsg.), *Nachhaltige Entwicklung* (S. 1–41). Springer Fachmedien Wiesbaden.
21. ISO (2017) ISO 22316:2017. Security and resilience—Organizational resilience—Principles and attributes. ISO. https://www.iso.org/standard/50053.html. Zugegriffen: 1. Sept. 2023.

22. Ivanov, D. (2020). Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. *Transportation Research Part E: Logistics and Transportation Review, 136,* 101922. https://doi.org/10.1016/j.tre.2020.101922.
23. Jeronen, E. (2013). Sustainability and sustainable development. In S. O., Idowu, N. Capaldi, L., Zu, A. D., Gupta (Hrsg.), *Encyclopedia of corporate social responsibility*, (S. 2370–2378). Springer Berlin Heidelberg.
25. Klasen, S. (2002). Low schooling for girls, slower growth for all? Cross-country evidence on the effect of gender inequality in education on economic development. *The world bank economic review, 16*(3), 345–373.
25. Modica, M., & Reggiani, A. (2015). Spatial economic resilience: overview and perspectives. *Networks and Spatial Economics, 15*(2015), 211–233. https://doi.org/10.1007/s11067-014-9261-7.
26. Rohde, F., Wagner, J., Reinhard, P., Petschow, U., Meyer, A., Voß, M., & Mollen, A. (2021). Nachhaltigkeitskriterien für künstliche Intelligenz, Bd 220.
27. Seeliger, L., & Turok, I. (2013). Towards sustainable cities: Extending resilience with insights from vulnerability and transition theory. *Sustainability, 5*(5), 2108–2128. https://doi.org/10.3390/su5052108.
28. Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature communications, 11*(1), 233. https://doi.org/10.1038/s41467-019-14108-y.
29. Vohra, S., Vasal, A., Roussiere, P., & Guan, L. (2022). The art of AI maturity: Advancing from practice to performance. Accenture Insights.
30. Wittenbrink, N., Demirci, S., & Wischmann, S. (2023). Resiliente und robuste KI-Systeme im praktischen Einsatz. In V., Wittpahl (Hrsg.), *Resilienz* (S. 199–211). Springer Berlin Heidelberg.
31. Yossi, S. (2005). *The resilient enterprise: overcoming vulnerability for competitive advantage*. The MIT Press. Pearson Education.
32. Zamani, E. D., Smyth, C., Gupta, S., & Dennehy, D. (2022). Artificial intelligence and big data analytics for supply chain resilience: A systematic literature review. *Annals of Operations Research,* 1–28. https://doi.org/10.1007/s10479-022-04983-y.
33. Zinke-Wehlmann, C., Friedrich, J., Kirschenbaum, A., Wölke, M., & Brückner, A. (2022). Conceptualizing sustainable artificial intelligence development. In L. M., Camarinha-Matos, A., Ortiz, X., Boucher, A. L., Osório (Hrsg.), *Collaborative Networks in Digitalization and Society 5.0* (Bd. 662, S. 545–554). Springer International Publishing.

Digital transformation of organizations or processes through AI

Abstract. The contributions to the session will highlight opportunities for the digital transformation of organizations and processes through the use of AI. The possible applications are diverse and range from the optimization of workflows and efficiency increases to the development of innovative business models based on AI methods. Specific real-world examples will be presented and discussed on how AI-based solutions can be used in different industries.

Developing a Human-centred AI-based System to Assist Sorting Laundry

Nico Rabethge(✉) and Franz Kummert

Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany
{nrabethge,franz}@techfak.uni-bielefeld.de

Abstract. This paper presents the development of a human-centred AI system for the classification of laundry according to washing categories such as color and type. The system aims to provide a solution that is both accurate and easy to use for individuals with varying levels of technical expertise. The development process involved a human-centred approach, including user research and testing, to ensure that the system meets the needs and expectations of its users. The system uses a combination of computer vision techniques and machine learning algorithms to analyze images of dirty laundry and provide recommendations for the appropriate washing category.

In addition to the development of the system itself, this paper also focuses on the explanation of the AI. The aim is to increase transparency and promote understanding of how the system makes its decisions. This is achieved through the use of visualizations and explanations that make the inner workings of the AI more accessible to users.

The results of testing demonstrate that the system is effective in accurately classifying dirty laundry. The explanation of the AI has yet to receive more feedback, whether users report that it increased their trust in the system and find it easy to use. The development of a human-centered AI system for laundry classification has the potential to improve the efficiency and accuracy of laundry sorting while also promoting understanding and trust in AI systems.

Zusammenfassung. In diesem Beitrag wird die Entwicklung eines menschenzentrierten KI-Systems für die Klassifizierung von Wäsche nach Waschkategorien wie Farbe und Typ vorgestellt. Das System zielt darauf ab, eine Lsung zu bieten, die sowohl einfach wie auch möglichst genau für Personen mit unterschiedlichem technischem Fachwissen zu bedienen sein soll.

Das System nutzt eine Kombination aus Computer-Vision-Techniken und Algorithmen des Deep Learning, um Bilder von schmutziger Wäsche zu analysieren und Empfehlungen für die richtige Waschkategorie zu geben. Neben der Entwicklung des Systems selbst geht es in diesem Beitrag auch um die Erklärung der KI und das Aktive Lernen. Ziel ist es, die Transparenz zu erhöhen und das Verstndnis dafür zu frdern, wie das System seine Entscheidungen trifft. Dies wird durch den

C. Zinke-Wehlmann und J. Friedrich (Hrsg.), *First Working Conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow*, AIDRST 2023, Informatik aktuell, S. 23–35, https://doi.org/10.1007/978-3-658-43705-3_3

Einsatz von Visualisierungen und Erklärungen erreicht, die den Nutzern die Funktionsweise der KI näher bringen. Durch das Aktive Lernen wird der Aufwand fr das Annotierten der Daten verringert, welches für jede Wscherei aufgrund unterschiedlicher Bedürfnisse erneut durchgeführt werden müsste.

Die Testergebnisse zeigen, dass das System in der Lage ist, bestimmte Attribute schmutziger Wäsche zuverlässig zu klassifizieren. Es sind zuknftig Nutzerstudien notwendig, welche überprfen, ob das Sytem tatschlich das Vertrauen in das System strkt und es einfach zu bedienen ist. Die Entwicklung eines menschenzentrierten KI-Systems zur Wäscheklassifizierung hat das Potenzial, die Effizienz und Genauigkeit der Wäschesortierung zu verbessern und gleichzeitig das Verstndnis und Vertrauen in KI-Systeme zu fördern.

Keywords: Computer Vision · Deep Learning · XAI · Human-centred approaches of AI design

1 Introduction

1.1 Problem

In industrial laundry technology, the sorting of dirty laundry according to washing categories has so far either not been carried out at all, has been done by scanning barcodes and RFID chips, or has been carried out manually by employees under safety measures. The latter requires human contact with the soiled and often contaminated laundry, which represents a great physical and psychological burden for the employees and also involves a considerable health risk. This problem arises in almost all industrial laundries (several hundred companies in Germany), so that support for laundries in this problem area would mean a huge improvement in working conditions for thousands and thousands of employees.

However, the application of AI in the laundry industry is currently very problematic. As in other industries, the term AI has extremely negative connotations because there is a general fear that AI will destroy jobs. Employees and their representatives are therefore skeptical about the introduction of AI. This is largely justified and must be part of a company's AI introduction strategy.

In some areas, however, the skepticism is also based on ignorance and misconceptions about AI and its practical applications. The actual goal of and the actual changes brought about by AI in the work process must be made clear. Educating employees through further training is thus a central prerequisite for the successful introduction of AI.

1.2 Objective and Solution Approach

The objective is to minimize the contact of working people with soiled and possibly contaminated laundry by automatically classifying the laundry delivered to the laundries so that it can be treated without direct human contact.

Based on an existing automatic machine for laundry separation, a solution for laundry sorting according to washing programs, primarily based on camera images and AI, is to be developed.

In addition, it must be taken into account that different types of laundry are washed in different laundries, so that a human-machine interaction module for intuitive training of the concrete requirements of a specific laundry shop will also be co-developed using "Active Learning" methods.

Moreover, the learned knowledge of the trained artificial neural network is not to function as a "black box", but rather current approaches of "Explainable Artificial Intelligence" are to be adapted and further developed to meet the needs of employees in medium-sized companies as well as increase their acceptance of AI.

This creates a new division of labor between humans and AI-controlled machines. This process of introducing AI control is to be supported by a qualification process that prepares and trains employees who are unaccustomed to cooperate with an AI system. The focus here is not only on technical training, but also on conveying a basic understanding of how this AI system works and the knowledge acquired, in order to minimize reservations about the use of this AI-based system.

Summarizing, the addition of a classification component to a wash chain of an industrial laundry shop improves working conditions in the area of laundry sorting. This is beneficial for the employees in industrial laundries and makes it easier for the mostly medium-sized companies to recruit workers for this not particularly popular area of work.

2 Related Work

2.1 Developing a Human-centred AI in an Industrial Setting

By taking over routine tasks (such as information search), the workload of humans (e.g., machine operators) can be reduced and value creation increased by focusing on core tasks. In the form of intelligent assistance systems, AI applications can support humans in specific tasks. To do this, the assistance systems analyze the current situation and make predictions. They are context-sensitive and can successively interact better with humans in a natural way (e.g., understanding speech) [11]. However, it is important to note that production conditions are subject to temporal variations, such as changing workloads and aging sensor systems. The data generated from this is not stationary, it is subject to drift, i.e. the observed data distribution is fundamentally different from the data used to train the systems. Efficient adaptive fusion architectures, where the learning process can be controlled by detected drift, are needed to facilitate human decision making in such problems while reducing complexity [4,5].

Furthermore, humans can adapt the AI, to minimize the drift. In order to accomplish this, however, it is necessary to make the AI's decisions understandable even to people with low technical expertise. In order to break the black-box character of modern AI models, a number of possibilities have been developed in recent years under the keyword "Explainable Artificial Intelligence" (XAI) to visualize the learned models of neural networks and to make them interpretable [12,17]. Thus, various methods of active learning (AL) have been developed for these networks, which, on the one hand, actively use human expertise in training these networks, and thus also reduce the necessary amount of hand annotation of huge datasets, which is particularly daunting for small- and medium-sized companies [13,16]. Through these methods, it is also possible to increase the understanding and therefore also trust of users.

2.2 Laundry Classification

Classification of textiles and cloth has been the subject of research in various fields, including computer vision, machine learning, and textile engineering. There are two popular open datasets, FashionMNIST and DeepFashion, both focusing on classifying the type of the clothing object. The pictures show the garments frontally and not bent/creased, thus in a totally different setting with less complexity compared to this project.

Due to the global increase in textile waste, various governments have issued orders for the recycling of textiles [19,20] and [21], increasing the research of classification of fabric. Kampouris et al. [10] used geometry to classify fabrics based on their surface textures and reflectance, while Sonawane et al. [15] employed a convolutional neural network (CNN) to recognize different types of clothing fabrics. Additionally, recent advances were able to achieve good classification rates based on hyperspectral data [3].

3 Methods

In this chapter, we will first discuss the conception of the project. Afterwards, the AI methods and those that support and simplify the work with the AI will be described.

3.1 Experimental Setup (AreaScan)

At first, a system was designed for a test bed, which is to be used for the recognition and classification of individual pieces of laundry. The concept can be seen in Fig. 1.

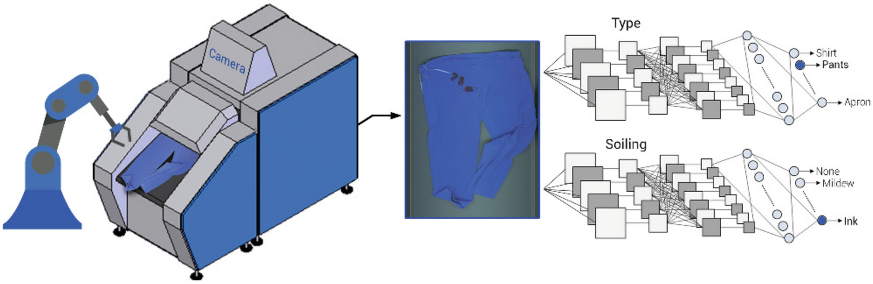

Fig. 1. Design of the identification process. The recognition module is shown on the left. When a piece of laundry passes over the conveyor belt, an image is stored. This is then processed by neural networks, which predict the individual attributes of the laundry item (right) to decide upon the correct washing category

In the middle of this module there is a gray conveyor belt over which the laundry pieces travel through the module one by one. The speed of the conveyor belt is 0.6m/s. A FRAMOS D415e depth camera with a resolution of 1280 X 720 pixels is mounted above the conveyor belt. In order to filter out variations in illumination, the conveyor belt surface is isolated from above, below and the sides by the walls and the conveyor belt itself. Lighting is provided by two LED strips in front of and behind the camera.

For the recording, the laundry pieces are separated and dropped at three different positions onto the conveyor belt. This is to simulate the dropping of a gripper arm, which in the future will take over the separation and dropping of the laundry pieces. The laundry then runs along the conveyor belt and reaches the camera's pickup area. When the camera detects a piece of laundry, several images are taken of the piece of laundry at regular intervals. This is necessary because some laundry items are too large for the recording area. A disadvantage of this methodology is that the laundry item is photographed from only one side and may also be scrunched up due to dropping. However, characteristics such as the color, type, and material should still be recognizable. In the case of soiling that is only visible on the underside, however, it is not possible to photograph it from above. Accordingly, the soiling would not be visible on the pictures. However, this restriction had to be made because in many laundries there is not enough space to hang up the laundry, so that it may be photographed from both sides. A possible solution to this problem could be a mechanism that turns the piece of laundry to the other side, whereupon another picture is taken.

In the next step, a manual annotation of the relevant classes is done. These categories are: color (white, black, red tones, blue tones, ...), type (shirt, trousers, gown, ...), degree of contamination (none, strongly, infectious, ...), kind of soiling (ink, mildew, blood), damage (none, chemical, mechanical) and washing temperature (30°, 60° and 90 (75)°). The classes are selected very finely. An adjustment for coarser classes for sorting is thereby also easily possible.

Adapted Setup (LineScan) When the module was exhibited and tested at a different location, the segmentation, and therefore the classification, was sometimes massively affected due to different lighting conditions despite the measures. Thus, the explanations and the observations we noticed during the annotation and segmentation process serve as a basis to make some optimizations to the recognition module.

The goal is to eliminate both the interfering objects and variance due to e.g. different lighting conditions. The interior should be shielded from the outside as much as possible, have a homogeneous/same color background and be well illuminated. Since we have laundry pieces of different lengths and depth data did not improve the classification, we decided to use a line scan camera as a reasonable adaptation. The line of the camera is aligned in such a way that it does not contain any interfering objects such as black blocks or holes. In addition, it is then only necessary to paint the walls in the area of the line to match the conveyor belt color and to adjust the illumination. Using this setup and using presorted laundry, two people were able to take and annotate $5,430$ pictures over 6 days, which is a speedup of factor 4.3 when considering the number of laundry items passing the camera compared to the first setup.

Due to feedback on our first working demonstrator and discussions with experts in the washing industry over time, we also adapted the categories. Color as well as soiling were extended by a few classes as well as considered as a multi-label classification problem. For the type, we only added a few new classes. We also added a new class "washing color", describing the color group with which the laundry item is to be washed together.

3.2 AI in Image Processing

We focus on the use of convolutional neural networks (*CNN*) [7], which are nowadays used in almost all areas of computer vision due to their great performance. These are artificial neural networks, which are particularly suitable for image data due to their design.

In order to simplify the handling of these methods for the employees, we also deal with methods of active learning and the explainability of artificial intelligence.

The way in which the neurons are arranged and connected is called the architecture of the artificial neural network. So far in this project, we have looked at architectures that achieved the best performance on the ImageNet dataset at the time of their publication [18]. For classification, the CNN architectures VGGNet [14] and DenseNet [6] were initially implemented to gain first insights. The basis for this decision was the low complexity of the VGGNet and the parameter efficiency of the DenseNet. For each of these architectures, one network per category was trained and optimized to keep the complexity as low as possible initially.

As we identified that the networks learned incorrect shortcuts for the classification (e.g. it has learned, that if the black blocks were visible at the edge, it has tended to be a small piece of laundry, which in almost all cases were colorful), we also trained PerturbGAN [1] and a U-Net [8] to extract the laundry object from the background. Therefore, only the area of the piece of laundry can be used for the classification.

3.3 Active Learning

The required annotation of this data adds to the effort, which would also have to be redone specifically for each different laundry shop due to different customers. Therefore, it makes sense to develop strategies that reduce the number of required annotated data or simplify the annotation. Such algorithms are assigned to the field of active learning.

The authors of the CEAL algorithm [16] (cost-effective active learning) propose to adopt the certain classifications obtained by the neural network on unlabeled data and use them in addition to the already annotated data for a new training of the same AI. This process is illustrated in Fig. 2. This process is repeated until either a satisfactory recognition performance of the neural network is achieved or the complete dataset is annotated. Such human-in-the-loop concepts can also increase acceptance and understanding.

3.4 Explainable AI

As already mentioned, the decision path of deep neural networks is difficult to understand. Artificial neural networks usually contain millions of parameters and links, which makes it very difficult for a human to keep track of them. This is why CNNs are also called black-box models. Explainable AI (*XAI*) methods try to shed light into this black box so that the neural network's decisions can be understood. In this project, we first looked at layer-wise relevance propagation (*LRP*) [2], as it produces an easy to understand heatmap in the input image indicating the level of importance of each pixel contributing to the final classification outcome.

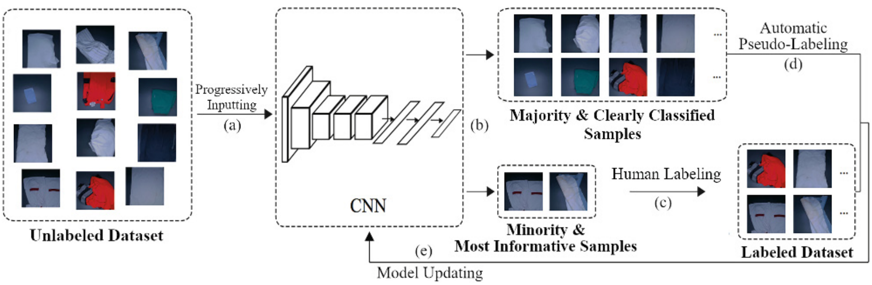

Fig. 2. The CEAL paradigm gradually feeds samples from the unlabeled dataset into a neural network (a). Then, the selection criteria for the certain images and the most informative/uncertain images are applied to the classification results of the neural network (b). After adding user-annotated uncertain images to the annotated dataset (c) and the pseudo-annotations of non-user-annotated certain images (d), the model is further updated/trained (e). Adapted from Wang et al. [16]

LRP considers the inputs and outputs of the network. Intuitively, it uses the network weights and the activations of the neurons generated by the forward pass to propagate the output back through the network to the input layer. This gives the input layer the relevance of each pixel to the predicted class. From this it can be deduced which pixels the network uses as positive (red), slightly positive (yellow) or neutral (green) indicators for the respective classification. Some LRP explanations in the context of this project can be seen in Fig. 3.

4 Results

We created two datasets, consisting of $9,405$ (AreaScan) and $5,430$ (LineScan) images. For each of the images, information exists about the color, the type, and mostly also about the soiling, the damage, and the material. For the AreaScan dataset, segmentation was considered before classification, so we present these results first below.

4.1 Segmentation

First, $5,000$ of the images were segmented by hand, as the unsupervised PerturbGAN did not provide good results. Subsequently, we iteratively trained a U-Net, predicted new segmentations for the remaining images and either accepted or rejected them until all images were segmented. Through this active learning oriented process, $2,091$ segmentations were automatically created. Without deducting the effort for accepting/rejecting, this saved about 22.23% (here specifically 32 hours) of work time. The segmentation results can be seen in Tab. 1.

Table 1. Accuracy and IoU on the test dataset of the best model for the AreaScan dataset. For comparison, we calculated the segmentation by means of difference to an image of an empty conveyor belt and the output of the GrabCut [9] algorithm without corrections

Score	Difference	GrabCut	PerturbGAN	U-Net
Accuracy	0.8971	0.9425	0.8864	0.9993
IoU	0.6989	0.7994	0.6106	0.9947

4.2 Classification

Dataset AreaScan In this dataset $6,855$ images (72.89%) are of laundry items that originate from nursing homes and $2,550$ images (27.11%) are of workwear. Since only for the color, the type and the soiling to more than 90% the classes could be determined with certainty and the dataset is already rather small in general, the focus was initially placed on these three categories.

For both CNN architectures, hyperparameter optimization was performed for each category combined with each dataset (unprocessed and segmented). We first restricted ourselves to the learning rate, batch size and to the use of a pre-trained network. Here, the best scores on the test dataset are listed in Tab. 2.

Table 2. Accuracy on the test dataset of the best model for the AreaScan dataset

Dataset	Color	Type	Soiling
AreaScan	0.9649	0.7989	0.7841
AreaScan (Segmentations)	0.9539	0.7952	0.8100

Dataset LineScan Here, using the 5,430 images, we also performed a hyperparameter optimization for both CNN architectures for each category. Since we now considered the classification of color as well as soiling from the creation of this dataset as a multi-label problem, we observed the F_1 score for these classes and had to change to a sigmoid activation function for the output. Furthermore, height equalization as preprocessing and data augmentation strategies like random rotation and vertical/horizontal flipping during training were performed. Here, the best scores on the test dataset are listed in Tab. 3.

Table 3. Accuracy (type, washing color) and F_1 score (color, soiling, damage) on the test dataset of the best model for the LineScan dataset

Dataset	Color	Type	Soiling	Washing Color	Damage
LineScan	0.8230	0.9652	0.5783	0.9713	0.8537

4.3 Explainable AI

In addition, the LRP algorithm has been implemented and can be applied to any network architecture. A visualization of the color classification explanations along with the certainty of the respective neural network can be seen in Fig. 3. Looking at the classification with high certainty (a-b, $> 98\%$), it is noticeable that only regions of the laundry piece were used for the classification. In the less certain range (c, $80 - 90\%$), the laundry piece was used for the most part, but also areas at the edge. In a relatively uncertain classification (d, $< 70\%$), the laundry piece was obviously not recognized and only edge pixels and pixels at the transition between conveyor belt and wall were used.

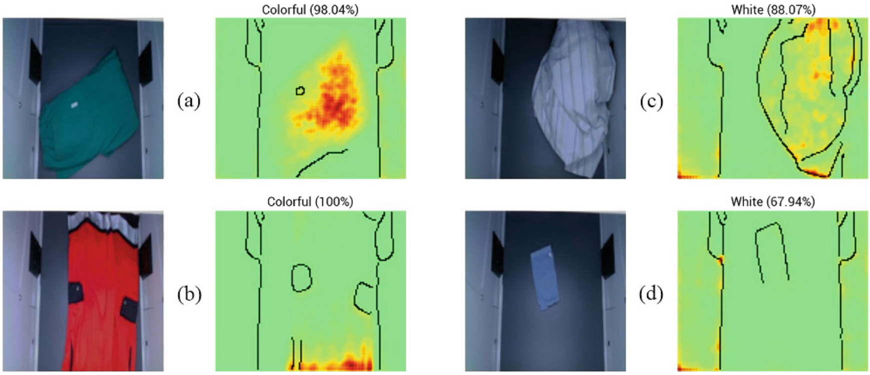

Fig. 3. Original AreaScan pictures (left) and the color classification including the LRP explanation (right). The relevance of each pixel for the classification (above) is color coded (green – neutral, red – very important)

4.4 Output of the System

Finally, Fig. 4 shows the representation of the whole output including certainties for each class and visualizations for a randomly chosen piece of laundry.

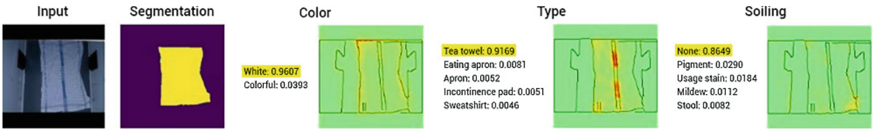

Fig. 4. Original images with the segmentations of the trained U-Net and the outputs of the most probable classes of a trained neural network per category. Here, the most probable class is highlighted in yellow. In addition, the relevance of the individual input pixels for the classification of the neural network is calculated using LRP and displayed in the form of a heatmap (green – neutral, red – very important)

5 Discussion

The CNNs were able to achieve good classification accuracies for all categories, although the AreaScan dataset being fairly small. By using LRP, unwanted strategies could already be detected, which is why networks were trained for segmentation. Since the unsupervised segmentation approach did not give good results, segmentations had to be created manually. Through the successful use of CEAL for creating the segmentations, time for the annotation process could be saved. The subsequent segmentation performance of the U-Net is very good. However, the classification could not be improved by using the segmented images. Nevertheless, the use of the shortcut could be eliminated.

Since, on the one hand, the laundry items were not completely visible on the images at the time and, on the other hand, tests revealed some optimization possibilities, the setup was adapted using a line scan camera, among other things. In addition, the recording process was further optimized, which quickly led to the creation of a second larger dataset (LineScan). On this one, the classification performances of the first trainings are for some classes already very good. In the next steps, CEAL will be used to further increase the size of the LineScan dataset and to continue training models. In parallel, further user studies are planned.

6 Conclusion

The paper proposes a human-centred system for sorting laundry based on deep learning models. The system uses convolutional neural networks to classify laundry items based on their visual features, and layer-wise relevance propagation to explain the model's predictions in a human-understandable way. The proposed system offers a promising approach for developing a user-friendly laundry sorting system which aims to take human needs and preferences into account, which will be examined in a subsequent study.

Acknowledgments. The results of this paper are based on the research project KISS within the Kompetenzzentrum Arbeitswelt.Plus. This research and development project is funded by the German Federal Ministry of Education and Research (BMBF) within "The Future of Value Creation – Research on Production, Services and Work" program and managed by the Project Management Agency Karlsruhe (PTKA). The author is responsible for the content of this publication. In cooperation with Herbert Kannegiesser GmbH.

References

1. Bielski, A., & Favaro, P. (2019). Emergence of object segmentation in perturbed generative models. *Advances in Neural Information Processing Systems,* 32. Annual Conference on Neural Information Processing Systems, pp. 7254–7264. NeurIPS 2019, Vancouver, BC, Canada.
2. Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R., & Samek, W. (2015). On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLOS ONE, 10*(7), 1–46.
3. Blanch-Perez-del-Notario, C., Saeys, W., & Lambrechts, A. (2019). Hyperspectral imaging for textile sorting in the visible-near infrared range. *Journal of Spectral Imaging, 8*(1), 1–11, Article ID a17.
4. Fritze, A., Mönks, U., Holst, C. A., & Lohweg, V. (2017). An Approach to automated fusion system design and adaptation. *Sensors 17*.
5. Holst, C. A., & Lohweg, V. (2018). A conflict-based drift detection and adaptation approach for multisensor information fusion. *In 23rd International Conference on Emerging Technologies and Factory Automation*, pp. 967–974. Turin, Italy.
6. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. (2017). Densely connected convolutional networks. In *IEEE conference on computer vision and pattern recognition*. https://doi.org/10.1109/CVPR.2017.243.

7. LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., & Jackel, L. D. (1989). Handwritten digit recognition with a back-propagation network. In D. S. Touretzky (Ed.), *NIPS* (pp. 396–404). Denver, Colorado, USA: Morgan Kaufmann.
8. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241. Springer.
9. Rother, C., Kolmogorov, V., & Blake, A. (2004). GrabCut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (SIGGRAPH), 23*(3), 309–314.
10. Kampouris, C., Zafeiriou, S., Ghosh, A., & Malassiotis, S. (2016). Fine-grained material classification using micro-geometry and reflectance. In *Computer Vision – ECCV 2016: Amsterdam, Netherlands*. Springer International Publishing.
11. Ludwig, B. (2015). *Planbasierte Mensch-Maschine-Interaktion in multimodalen Assistenzsystemen* (1st ed.). Springer.
12. Samek, W., Wiegand, T., & Müller, K. R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. ITU Journal: ICT Discoveries – Special Issue 1 – The Impact of Artificial Intelligence (AI) on Communication Networks and Services 1, 1–10.
13. Sener, O., & Savarese, S. (2018). Active learning for convolutional neural networks: A core-set approach. In: ICLR (Poster). OpenReview.net, Vancouver, BC, Canada.
14. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv 1409.1556*.
15. Sonawane, C., Singh, D. P., Sharma, R., Nigam, A., & Bhavsar, A. (2019). Fabric classification and matching using CNN and siamese network for E-commerce. In *Computer Analysis of Images and Patterns*, pp. 193–205. Springer International Publishing.
16. Wang, K., Zhang, D., Li, Y., Zhang, R., & Lin, L. (2016). Cost-effective active learning for deep image classification. *IEEE Transactions on Circuits and Systems for Video Technology, 27*, 1–1.
17. Zhang, Q., & Zhu, S. (2018). Visual interpretability for deep learning: a survey. *Frontiers of Information Technology & Electronic Engineering, 19*(1), 27–39.
18. Papers With Code ImageNet Ranking, http://www.'aperswithcode.com/sota/image-classification-on-imagenet'. Last accessed 27 March 2023.
19. Chinese Government Homepage on textile recycling, http://www.'english.www.gov.cn/statecouncil/ministries/202204/20/contentWS625f649fc6d02e5335329a8f.html'. Last accessed 27 March 2023.
20. Australian Government Homepage on the textile waste roundtable, http://www.'www.dcceew.gov.au/environment/protection/waste/product-stewardship/textile-waste-roundtable'. Last accessed 27 March 2023.
21. European Commission Homepage on the EU strategy for Sustainable and Circular Textiles, http://www.'ec.europa.eu/commission/presscorner/detail/en/QANDA222015'. Last accessed 27 March 2023.

AI-Powered Knowledge and Expertise Mining in Healthcare from a Field Experiment

Julia Kauper[1(✉)], Susanne Franke[2], Felix Franke[2] and Steven Grieshammer[3]

[1] Staatliche Studienakademie Plauen, Schloßberg 1, 08523 Plauen, Germany
julia.kauper@ba-sachsen.de
[2] d-opt GmbH, Oberneumarker Str. 59, 08496 Neumark, Germany
s.franke@d-opt.de
[3] Heinrich-Braun-Klinikum gGmbH, Karl-Keil-Str. 35, 08060 Zwickau, Germany

Abstract. With the increasing prevalence of mobile applications across various domains, there is a growing demand for individualised and self-adaptive learning pathways. This is particularly important in the mobile health sector, where there is a critical need to investigate how expert and experiential knowledge can be acquired, digitalised and formalised into data which is subsequently processed and further used. To address this demand, our research explores how Artificial Intelligence (AI) can power this process. We developed a prototype mobile application with a standardised learning pathway that features speech-language therapy exercises of varying levels of difficulty. In a 12-week field experiment involving 21 individuals with aphasia, we analysed the results using supervised and unsupervised algorithms. Our findings suggest that AI has the potential to generate new knowledge, such as identifying features that can determine which learning words are perceived as easier or more difficult on an inter-individual basis. This knowledge enables algorithmisation and the design of standardised (database-supported) artefacts, which in turn can be used to formulate self-adaptive and individualised learning pathways. This significantly enhances the development of effective mobile applications to assist speech-language therapy.

Keywords: Artificial Intelligence · Expertise Mining · Mobile Health · Speech-Language Therapy

1 Introduction

As the world becomes increasingly digitalised, there is a growing emphasis on the human element within technological environments. Technology is being used to adapt processes in e.g., production, healthcare, and education to better serve human needs [1]. With the shift from Industry 4.0 to Industry 5.0, the focus is set on promoting human well-being and developing skills and abilities [2]. Artificial Intelligence (AI) is an important and rapidly evolving technology that can support this process. Qualification processes for humans are of particular interest and relevance, for example, in training new employees and reskilling workers in the context of work, as well as in re-

C. Zinke-Wehlmann und J. Friedrich (Hrsg.): *First Working Conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow*, AIDRST 2023, Informatik aktuell, S. 37–49, 2024. https://doi.org/10.1007/978-3-658-43705-3_4

covery from serious illnesses or medical procedures. In all these cases, there is a need for individualised learning processes that are tailored to each individual's skill level and any relevant restrictions (e.g., health or knowledge). Our goal is to formulate individualised learning content and exercises to optimise the learning progress for each participant. In this paper, we investigate how AI can support this process by providing deeper insights into the available data base. The paper is structured as follows: Sect. 2 provides background information, Sects. 3 presents the proposed methods, Sect. 4 and 5 describe the experimental results, and Sect. 6 provides a conclusion and outlook for future work.

2 Theoretical Background and Current State of the Art

AI as a branch of computer science deals with the development of algorithms to perform tasks traditionally associated with human intelligence, like the ability to learn and solve problems. A thorough statistical basis and an introduction into Machine Learning approaches is given in [3]. In [4], the focus is set on the algorithmic viewpoint and e.g. complexity analysis, whereas [5] provides an introduction into Deep Learning. AI applications are already prevalent in diagnostics, especially in the field of computer vision for tumour identification, see e.g., the study [6]. Recent developments were achieved in disease classification based on segmentation [7] and in natural language processing [8].

Especially in the context of Industry 5.0, the focus is set on the "human in-the-loop" (HITL), whereas AI is seen as supporting technology: The human provides an extensive data base of high quality (which is required to not be biased), detects poor training results, interprets them and extracts knowledge for further use. Explainable AI for methods which do not provide such an explanation of the solution-finding process (as e.g. the decision tree algorithm does) is an avid field of research. For an overview, see [9]. However, the results are mostly theoretical and not yet in application. The choice of an algorithm highly depends on the planned usage of the results. From a practical point of view, users prefer to apply algorithms that provide a high level of (inherent) explainability. Measuring learning progress and learning analytics is an active field of research, as evidenced by numerous publications. A reference model is given by [10], whereas [11] provides a framework of quality indicators to standardise the evaluation of learning analytics tools. Various empirical studies and implementations, such as [12], also exist.

HITL and AI-enabled adaptive learning systems have several capabilities such as generating individualised learning pathways adapted to individual progress, preferences, and pace [13]. A user-centred approach enables system feedback and correction, adapting to users' learning styles and knowledge, providing individualised learning pathways aligned with specific needs and individual goals [14], ensuring higher accuracy and effectiveness by adapting to changing conditions and data [13]. Adaptive learning systems have broad applications across domains, including language education using frameworks such as the Common European Framework of Reference for Languages (CEFR) [15]. Speech-language therapy lacks such frameworks and relies on the expert and experiential knowledge of speech-language therapists (SLTs) using different treatment guidelines, techniques, and technologies based on the principles of evidence-based practice (EBP), comprehensive diagnosis, and individual therapy goals [16].

Digital applications used in speech-language therapy include high-tech augmentative and alternative communication (AAC) tools such as speech-generating devices, apps on tablets [17] and interactive tables with touchscreens [18]. There is also an increasing usage of mobile apps [19]. Machine learning can enable SLTs to automatically adjust training and rehabilitation devices remotely based on a tailored treatment plan [20], thereby providing a personalised and more effective therapy experience. Hence, there is a growing need to explore expert and experiential knowledge in data to design self-adaptive and individualised learning pathways.

3 Methodology

To create self-adaptive and individualised learning pathways for mobile applications in speech-language therapy, expert and experiential knowledge needs to be translated into data and implemented as artefacts in an app for testing and evaluation. Following an iterative approach, we combine two different methods to achieve this goal: 1) the enhanced Action Design Research (ADR) process for the design and development of the mobile app, and 2) the Cross-Industry Standard Process for Data Mining (CRISP-DM) as part of the evaluation to gain domain knowledge and thus improve the artefacts.

ADR aims to tackle a real-world problem within a specific organisational context while simultaneously designing and evaluating an IT artefact that deals with a set of issues inherent to that situation [21]. By incorporating a field experiment into the ADR process for iterative refinement and evaluation of the mobile app, we combined design science and field experiments as described in [22].

CRISP-DM is an open, industry-, software- and application-independent standard process model and provides an approach for data mining processes. The model consists of six phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Assessment and Deployment. [23]

The following sections describe the application and results of these methods in more detail.

4 Enhanced ADR and Field Experiment

The app development and the associated transformation of expert and experiential knowledge into data were guided by the enhanced ADR with the following stages:

Stage 1) Problem formulation: To understand how a mobile app can support speech-language therapy, patients' specific needs and objectives were identified following EBP as recommended by the professional guidelines for speech therapy of the German Federal Association for Speech Therapy (dbl) [24], taking into account factors such as patient expectations, therapist's clinical experience, and available scientific evidence [25]. Data was collected through scientific research and interviews with SLTs to investigate feasibility, including information on patient diagnoses and individual therapy goals. Therapy goals take into account factors like speech impairment type and severity, individual needs and preferences, and participation limitations according to the International Classification of Functioning, Disability and Health ICF [26]. Despite challenges in achieving standardisation, considerations were made towards standardisation and algorithmisation in designing a standardised database-supported portfolio that considers individual needs, the severity and manifestation of speech disorders, and different therapy goals.

Stage 2) Building, intervention, and evaluation: SLTs designed a uniform core learning pathway for all participants as a foundation for creating individualised learning pathways based on therapy goals, diagnoses, patient-specific needs and abilities, and app usage data. This core learning pathway consisted of a predefined sequence of three evidence-based speech-language exercises per day on at least five days per week such as picture-word association, spelling, and selection of the correct sentence with increasing difficulty. Semantic fields (categories) with medium to high intersubjective relevance to daily life of the participants, including food, clothing, housing, body, everyday objects, garden, illness, and personal care, taking into account the gender and age distribution of the patients, were identified. Subsequently, we selected related target words, defined data requirements for the prototypical mobile app for speech-language therapy, and evaluated user acceptance, design, as well as usability through field trials.

Stage 3) Field experiment: A 12-week field experiment was conducted from January to April 2022 with 21 post-stroke aphasia patients in Germany with varying levels of severity to assess the efficacy and usability of the app under real-world conditions. Post-stroke aphasia is a language disorder, affecting any aspect including speaking, writing, reading, and comprehension, that arises following a stroke [27] and is often accompanied by comorbidities such as motor impairments like hemiparesis and cognitive deficits like memory loss [28]. Eligibility criteria included a diagnosis of aphasia, premorbid competence in German, and sufficient auditory/visual acuity to interact with a tablet. Some eligible participants were excluded due to resource constraints, i.e., the lack of access to a device or internet. Participants' mean age was 63.6 years (range 38–81), and 67% were male. Personal information was provided through an anamnesis questionnaire, including visual and hearing impairment, handedness, and stroke history. Data on the specific type of aphasia, severity, type of stroke and co-morbidities were not collected as they were not considered relevant to the inclusion or exclusion criteria. This is because the design of the speech-language exercises aimed to take into account the heterogeneity of clinical disorders observed in people with aphasia. User feedback was collected through two questionnaires: a usability test at first use and a final questionnaire, which included self-assessment of the difficulty of the exercises. Progress was monitored, analysed, and evaluated at both individual and inter-individual level.

Stage 4) Reflection and learning: Through the evaluation of the field experiment's results, we engaged in deliberate and conscious reflection on the problem formulation, our working hypotheses, and the app artefacts that were developed. The next section presents and discusses the outcomes of the first five phases of the CRISP-DM deployment process. These results will be applied in *Stage 5* of the enhanced ADR, *Formalisation of learning*, to optimise and formalise our models and assumptions.

5 Exploration of Expert Knowledge Using CRISP-DM

To optimise the learning curve for each individual patient, the exercises have to contain words of a suitable level of difficulty. In the following, data-based evidence is extracted to evaluate the chosen level of difficulty for each word and to provide relevant features, extract potentially irrelevant features and suggest additional features to be acquired. Therefore, we applied CRISP-DM to the example of the picture-word association exercise and its difficulty factors to answer the following two questions: (1) How do the assumptions on difficulty-determining features affect the frequency of errors per word across users with heterogeneous clinical presentations in terms of "easy", "medium" or "hard"? (2) Which features lead to the classification of a target word as "easy", "medium" or "hard"?

5.1 Understanding of the Use Case

The picture-word association exercise in the app prototype presents a picture of the target word on the right side of the screen and four written words (the target word and three distractors) on the left and centre. Participants must correctly identify and match the word with the corresponding picture. The exercise's difficulty increased weekly based on factors such as the type of exercise, prototypicality and unambiguous representation of the target word in the picture, semantic field (category), and, more detailed below, target word/item and distractors.

Several factors determine the difficulty of the target word. These include individual frequency of use [29], degree of concreteness according to the Dual Coding Theory (DCT) [30], prototypicality (e.g., "Möhre" (carrot Eng.) is more prototypical than "Schwarzwurzel" (salsify Eng.) for the category *vegetables*), length and structure of the word (shorter words with simpler syllable structures (e.g., CV) are easier to learn than longer words with more syllables and complex syllable structures [31]), level of specification within a semantic hierarchy (e.g., "Hund" (dog Eng.) is easier to learn than "Dackel" (dachshund Eng.)), and properties such as dialectal peculiarities/synonyms, imageability, number of phonemes, compound words, and foreign words.

The properties of the three distractors (semantic, phonological, and non-relational) in relation to the target item also influence the difficulty level of this exercise. For *semantic distractors*, the level of difficulty increases with the closeness to the target item (equivalence in terms of prototypicality/type frequency, imageability/visual redundancy or distinctiveness, semantic/relational proximity). For example, "Keks" (cookie Eng.) is an easier semantic distractor for the target word "Möhre" in the semantic field food than "Rettich" (radish Eng.). For *phonological distractors*, it holds that the more similar the phonological-lexical properties (compared to the target word), the more difficult this distractor is [32]. For example, "Matratze" (mattress

Eng.) has the same initial letter as “Möhre”, but a different number of syllables as well as a different number of graphemes and is therefore an easy phonological distractor. “Möwe” (seagull Eng.) is a more difficult distractor, as it has the same initial letter, same number of syllables, similar syllable structure, same vowel length, and it also rhymes with “Möhre”. *Non-relational distractors* must not be in a semantic or phonological-lexical relation to the target word or to any of the other distractors, like “Goldkette” (gold chain Eng.) for “Möhre” (carrot Eng.), as frequent errors with this distractor indicate guessing.

Data collection took place during and after the 12-week test phase and included questionnaires, recorded app usage data, a vocabulary database with items (e.g., target words, and distractors) created by the SLTs, and developed attributes. The criteria that determine the level of difficulty of an item were elaborated and described in terms of their characteristics.

5.2 Data Understanding

This phase involved a thorough investigation of the provided database. The data on targets included around 25 features, with 11 having been identified as relevant for this investigation by the SLTs and data analysts. Hereby, certain features were neglected for this investigation (this mainly depended on the availability and validity of the data). Since the test phase was already completed, it wasn’t possible to acquire additional data from the test phase itself as well as patient-specific characteristics (e.g., displayed distractors which weren’t clicked; the patients’ familiarity with items from certain semantic fields).

The selected features included categorical values like difficulty and concreteness/abstractness, integer values like number of graphemes, Boolean values like compound words, and strings like structure syllables. The same attributes were provided for the distractors, with the additional feature “type” with its feature space {semantic, phonological, non-relational}. Additionally, data from the test phase was provided, including e.g., exercise ID, patient ID, time stamps of each click within the exercise-solving process by the patient, correct target value, falsely chosen distractors, used support (audio, joker). For a better understanding of the data, uni- and multivariate data analysis was performed: Histograms and boxplots allowed a deeper insight on distributions and possible outliers (which occurred, for example, regarding the duration of an exercise or the total number of errors per item). Additionally, the calculation of correlation coefficients and category-wise variance analysis provided a deeper understanding of the features’ connections (e.g., there is an obvious correlation between the length of a word and the number of its syllables, but the word type frequency is not correlated with the other item’s features and the type-specific number of errors). Appropriate visualisations, both individual and inter-individual, included e.g., the number of trainings per day and exercise (for each week) as stacked bar charts, the distribution of the types of errors (semantic, phonological, and non-relational) as pie charts as well as box plots, and the temporal distribution of the trainings as scatter plots.

5.3 Data Preparation

In the context of univariate data analysis, it became apparent that the duration of some picture-word association exercises lasted up to 23 times as long as the average time and/or were not finished. Unfinished exercises were neglected; regarding the duration, a time limit was set in coordination with the SLTs. Furthermore, plausibility checks performed in Python 3.9 showed that for exercises exceeding one day, incorrect data was stored in the database regarding the type of exercise. These mistakes were corrected. There were no missing values.

Further, it was required to integrate the data from the different sources. The characteristics of the targets and distractors were provided in various Microsoft Excel files, as they were assembled manually by the SLTs. Data from the test phase were acquired automatically and made available in a MySQL database. The relevant data was extracted and structured with regard to the analysis to be performed in a new MySQL database. This also included the calculation of new features like the type-specific (semantic, phonological, and non-relational) number of errors per exercise.

After data cleaning and integration, the data was then prepared in accordance with the AI algorithms to be applied. For the decision tree algorithm, both numerical and categorical data are allowed, and the data do not need to be normalised. Regarding clustering methods, the preprocessing process included distribution and outlier handling, transformation and normalisation of numerical data, and removing highly correlated features.

5.4 Modelling

Two different AI approaches were implemented. First, the decision tree as supervised method was applied to identify the features and their values leading to the difficulty of each target (with difficulty space {easy, medium, hard}). The main advantage of this approach were the explainable results, which means that the features and their respective values leading to the categorisation could be provided and discussed with the SLTs. As hyperparameters, the gini criterion was chosen, with a maximum depth of 10 being sufficient to provide unique classification results.

Second, clustering algorithms were applied as unsupervised methods to check which grouping of items was suggested. For centroid-based methods, the elbow method to identify the optimal number of clusters led to three (which matched the number of categories the SLTs chose). For comparison, two and four clusters were also set as hyperparameters. Additionally, connectivity-based methods were applied.

Besides the items' characteristics (which formed the basis for the SLTs to derive the difficulty of the items), the results from the test phase could also be included, i.e., the number of mistakes for each of the three difficulty types. Hereby, the types semantic, phonological, and non-relational distractors were distinguished. Both approaches (leaving out and including data from the test phase) were applied and evaluated individually.

Additionally, statistical methods for multivariate analysis were applied to compare the feature values of targets with same difficulty. This allowed the identification of conspicuous targets or peculiarities with respect to the semantic fields. For example,

Fig. 1 in Sect. 5.5 shows the comparison of selected feature values of target words with the same difficulty for the semantic fields clothing and food.

5.5 Evaluation

The results were discussed with the SLTs. Regarding the decision tree, an iterative investigation of the leaf nodes revealed items which weren't categorised easily and required deep branching. Targets which were difficult to categorise had an 85% intersection for both approaches (leaving out and including test phase data), indicating a high consistency of the SLTs' and patients' perception. If considered, the relative number of errors heavily influenced the decision nodes, but not the results as much – especially if a sufficiently high depth was chosen as hyperparameter. Tab. 1 shows the impact of features selected for the algorithm with maximum depth value of 5. Here, the relative number of misclassified targets (also specifying the respective falsely stated difficulty) is depicted for three cases of selected features. It becomes obvious that including the data acquired during the test phase with the patients significantly improves the results. Simultaneously, the typical problem of words misclassified as "medium" is still apparent.

Tab. 1. Number of misclassified targets under different features selected for depth 5

Decision tree of depth 5	Easy	Medium	Hard
Features excluding errors			
Misclassified as "Easy"	–	10%	17%
Misclassified as "Medium"	17%	–	17%
Misclassified as "Hard"	3%	8%	–
Features including errors in total			
Misclassified as "Easy"	–	12%	3%
Misclassified as "Medium"	3%	–	17%
Misclassified as "Hard"	3%	2%	–
Features including all types of errors			
Misclassified as "Easy"	–	2%	3%
Misclassified as "Medium"	0%	–	17%
Misclassified as "Hard"	0%	0%	–

Five features were identified to have major impact on the categorisation. Certain semantic fields could be classified easily as suggested by the SLTs, whereas targets with "medium" difficulty were the most problematic ones. These items were further investigated using clustering results: Changing the number of clusters (and, hence, the number of difficulty categories) to two provided a deeper insight into the connection between the item characteristics and the results from the test phase. Most of the target values with difficulty "easy" and "hard" were assigned to the two different clusters, respectively, indicating that the SLTs' classification of "easy" and "hard" was mostly in accordance with patients' perception. Target values with "medium" difficulty were reassorted. It became apparent that for each semantic field with this difficulty, either a reassignment to "easy" or "hard" heavily predominated. The results brought deeper insight into which features might be relevant for the assignment to a cluster. Creating four clusters is however challenging from the application point of view: A further division of the difficulty "medium" would be necessary. As for three clusters, about half of the targets matched the SLTs' classification, where the targets identified as critical with the decision tree method were assigned to different clusters with this method as well. These results in connection with the target analysis on targets with same difficulty provided the basis for the SLTs to find causes for targets with a high deviation from the original difficulty. Such reasons were, e.g., an unsuitable or ambiguous picture provided in the test phase.

Several working hypotheses were confirmed by cross-patient valid attribute rules. For example, the difficulty of exercises was found to depend on the type of exercise, and the percentage of phonological and semantic errors in the picture-word association exercise increased with higher difficulty, indicating appropriate distractors were developed for these levels of difficulty. Additionally, for the category food at an easy level, there was a positive correlation between word type frequency and error rate. This confirmed the hypothesis that words that occur more frequently in the German corpus (e.g., "Marmelade", jam Eng.) would elicit fewer errors than words that occur less frequently (e.g., "Käse", cheese Eng., or "Tee", tea Eng.). However, word type frequency alone was not sufficient to explain the error patterns. Type frequency from the lexical database dlexDB [33] was used, which is based on the reference corpus of the German language of the 20th century from the Digital Dictionary of the German Language (DWDS) and ranges from 1900 to 1999. Therefore, the type frequency of certain words does not allow any conclusion on the current usage frequency of these words. For instance, we found more errors for words with shorter word length than for words with longer word length within the same category and level of difficulty. This contradicted the assumption regarding word type frequency and number of syllables. Possible factors that could account for this discrepancy are syllable structure, picture quality, distractor choice, or everyday relevance (depending on each patient's perception). We also performed pair- and groupwise comparisons. For the two categories food and clothing at an easy level of difficulty, the similarity between features like word type frequency across both categories is depicted in Fig. 1. However, there were fewer phonological and semantic errors for food compared to clothing. A possible conclusion is that the visual representation of food items may show a greater degree of prototypicality and a closer alignment with everyday visual perception as compared to items of clothing. Furthermore, food might activate other sensory modalities such as taste.

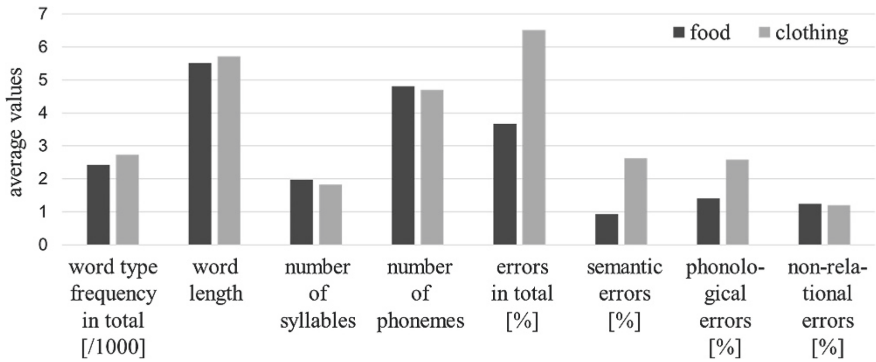

Fig. 1. Comparison of the selected feature values of target words with the same level of difficulty for the categories clothing and food

6 Conclusion and Outlook

Research on the transformation and exploration of expert and experiential knowledge into data using AI is crucial for the development of self-adaptive and individualised learning pathways in mobile applications to support patient-centred speech therapy treatments. In this study, we developed a prototype app, tested it with aphasic patients in a field experiment, and subsequently analysed and evaluated the results using different AI approaches. The study confirmed some of the assumptions and criteria made by SLTs for developing a standardised core learning pathway, while also uncovering patterns and weaknesses that can enhance the revision and improvement of the prototype.

Further development of the prototype could enable patients with speech disorders to train continuously with increasing demands under professional supervision. The individual therapy goal setting, taking into account participation goals (ICF) and inter-individual comparisons using newly developed algorithms for comparability and pattern recognition, is a pioneering basis. The significance of the current results is mainly restricted by the small number of patients and the brevity of the test phase. While the application of the decision tree algorithm allows a high transparency for the SLTs, it is highly susceptible to overfitting. Systematically selecting different features and applying suitable maximum depths helped to avoid those downfalls; with a bigger data set, random forests as an ensemble method could also be applied. In the future, we plan to test Deep Learning approaches as well to get new insights. The inherent problem of decision trees to be biased towards the dominant class in imbalanced data sets must be considered as well. Clustering methods, on the other hand, as unsupervised methods, have the disadvantage of not providing explainable results. Additionally, the quality of the results highly depends on the data points' location in the feature space, such that suitable algorithms must be chosen.

To overcome the described inaccuracies, we plan to increase the sample size for future studies, adjust the app and vocabulary databases, and revise the criteria for determining the difficulty of a target word and exercise. Additionally, a more precise description of individual and inter-individual evaluations would enable a more accurate definition of the data collection requirements, including additional patient-specific data. We also recommend developing new item-specific attributes that can adapt the difficulty of an exercise to individual learning progress.

Acknowledgments. This work was partly funded by the EFRE/ESF SMS funding program "Förderung innovativer Ansätze im Bereich der Gesundheits- und Pflegewirtschaft", project IMPULS-SACHSEN 2.0, and co-financed with tax funds on the basis of the budget approved by the Saxon State Parliament, project Sprach-Impulse. We would like to express our gratitude to all those who have contributed to this paper in various ways. We would like to thank the study participants for their time and cooperation, and all project partners for sharing ideas and data.

References

1. European Commission, Directorate-General for Research and Innovation, Breque, M., De Nul, L., Petridis, A.: Industry 5.0: towards a sustainable, human-centric and resilient European industry, Publications Office (2021). https://data.europa.eu/doi/10.2777/308407.
2. Rannertshauser, P., Kessler, M., & Arlinghaus, J. C. (2022). Human-centricity in the design of production planning and control systems: A first approach towards Industry 5.0, *IFAC-PapersOnLine 55*(10), 2641–2646. https://doi.org/10.1016/j.ifacol.2022.10.108.
3. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning* (1st edn., pp. 67–227). New York: Springer.
4. Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning. From Theory to Algorithms* (1st edn., pp. 115–284). Cambridge University Press.
5. Goodfellow, 1., Bengio, Y., & Courville, A. (2016). *Deep Learning* (pp. 162–481). The MIT Press.
6. Dembrower, K., Wåhlin, E., Liu, Y., Salim, M., Smith, K., Lindholm, P., Eklund, M., & Strand, F. (2020). Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. *Lancet Digit Health, 2*(9), 468–474. https://doi.org/10.1016/S2589-7500(20)30185-0.
7. Kok, E.Y., Pszczolkowski, S., Law, Z. E., Ali, A., Krishnan, K., Bath, P. M., ... & French, A. P. (2022). Semantic Segmentation of Spontaneous Intracerebral Hemorrhage, Intraventricular Hemorrhage, and Associated Edema on CT Images Using Deep Learning. *Radiology: Artificial Intelligence, 4*(6),3–4. https://doi.org/10.1148/ryai.220096.
8. Fink, M. A., Kades, K., Bischoff, A., Moll, M., Schnell, M., Küchler, M., ... & Kleesiek, J. (2022). Deep Learning-based Assessment of Oncologic Outcomes from Natural Language Processing of Structured Radiology Reports Radiology. *Artificial Intelligence, 4*(5),5–7. https://doi.org/10.1148/ryai.220055.
9. Gohel, P., Singh, P., & Mohanty, M. (2021). Explainable AI: Current status and future directions. *arXiv preprint* arXiv:2107.07045. https://doi.org/10.48550/arXiv.2107.07045
10. Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2013). A reference model for learning analytics. *International Journal of Technology Enhanced Learning, 4*(5–6), 318–331.
11. Scheffel, M., Drachsler, H., Stoyanov, S., & Specht, M. (2014). Quality indicators for learning analytics. *Educational Technology & Society, 17*(4), 117–132.

12. Ifenthaler, D., & Yau, J. Y. K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Education Technology Research and Development 68,* 1961–1990. https://doi.org/10.1007/s11423-020-09788-z.
13. Kabudi, T., Pappas, I., Olsen, D. H. (2021). AI-enabled adaptive learning systems: A systematic mapping of the literature. *Computers in Education: Artificial Intelligence, 2,* 100017. https://doi.org/10.1016/j.caeai.2021.100017.
14. Vesin, B., Mangaroska, K., & Giannakos, M. (2018). Learning in smart environments: user-centered design and analytics of an adaptive learning system. *Smart Learning Environments, 5*(1), 24. https://doi.org/10.1186/s40561-018-0071-0
15. Council of Europe. (2020). Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume, Council of Europe Publishing, Strasbourg, available at www.coe.int/lang-cefr, Accessed: 26. March 2023.
16. Grötzbach, H. (2018). Evidenzbasierte Praxis als Grundlage sprachtherapeutischer Entscheidungen. *Sprache-Stimme-Gehör, 42*(03), 133–139.
17. American Speech-Language-Hearing Association (ASHA): Augmentative and Alternative Communication (AAC) Homepage, http://www.asha.org/public/speech/disorders/AAC/, Accessed: 15. March 2023.
18. Plößel, L., Heide, J., & Netzebandt, J. (2021). Einsatzmöglichkeiten des ActivTable in der kognitiv orientierten Sprachtherapie: eine Exploration. In: Fritzsche, T., Breitenstein, S., Wunderlich, H., Ferchland, L. (eds.): *Spektrum Patholinguistik* (Band 14., pp. 83–92). Universitätsverlag Potsdam, Potsdam. https://doi.org/10.25932/publishup-51034.
19. Meier, B., & Leinweber, J. (2023). Ein Update zu digitalen Interventionen in der Logopädie/Sprachtherapie. *Neurologie & Rehabilitation, 29*(1), 27–33.
20. Mulfari, D., La Placa, D., Rovito, C., Celesti, A., & Villari, M. (2022). Deep learning applications in telerehabilitation speech therapy scenarios. *Computers in Biology and Medicine, 148*, 105864. https://doi.org/10.1016/j.compbiomed.2022.105864.
21. Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action Design Research. *MIS Quarterly, 35*, 37–56. https://doi.org/10.2307/23043488.
22. Sun, R. T., Han, W., Chang, H. L., & Shaw, M. J. (2021). Motivating adherence to exercise plans through a personalized mobile health app: enhanced action design research approach. *JMIR mHealth and uHealth, 9*(6), e19941. https://doi.org/10.2196/19941.
23. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide, Article 13.
24. Deutscher Bundesverband für Logopädie e.V. (dbl) (2012). Berufsleitlinien Logopädie. Fassung des Beschlusses der Mitgliederversammlung vom 04.06.2010. Frechen.
25. Aichert, I., & Staiger, A. (2015). Therapie der Sprechapraxie. *Forum Logopadie., 29*(5), 30–37. https://doi.org/10.1055/b-0037-144279.
26. World Health Organization (WHO): International Classification of Functioning, Disability and Health (ICF). https://www.who.int/classifications/international-classification-of-functioning-disability-and-health, Accessed: 3. March 2023.
27. Berthier, M. L., & Pulvermüller, F. (2011). Neuroscience insights improve neurorehabilitation of poststroke aphasia. *Nature Reviews Neurology, 7*(2), 86–97.
28. Ogar, J., Slama, H., Dronkers, N., Amici, S., & Gorno-Tempini, M. L. (2005). Apraxia of speech: An overview. *Neurocase, 11*(6), 427–432.
29. Cherney, L. R., Patterson, J. P., Raymer, A., Frymark, T., & Schooling, T. (2008). Evidence-based systematic review: Effects of intensity of treatment and constraint-induced language therapy for individuals with stroke-induced aphasia. *Journal of speech, language, and hearing research 2008 Oct;51*(5), 1282–99. https://doi.org/10.1044/1092-4388(2008/07-0206).
30. Paivio, A. (2014). Intelligence, dual coding theory, and the brain. *Intelligence, 47*, 141–158. https://doi.org/10.1016/j.intell.2014.09.002.

31. Nickels, L., & Howard, D. (2004). Dissociating effects of number of phonemes, number of syllables, and syllabic complexity on word production in aphasia: It's the number of phonemes that counts. *Cognitive Neuropsychology, 21*, 57–78.
32. Bi, Y., Xu, Y., & Caramazza, A. (2009). Orthographic and phonological effects in the picture–word interference paradigm: Evidence from a logographic language. *Applied Psycholinguistics, 30*(4), 637–658. https://doi.org/10.1017/S0142716409990051.
33. dlexDB Dokumentation. http://www.dlexdb.de/documentation/overview/, Accessed: 21. March 2023.

Iterative Development of a Process-Oriented Approach for the Selection of Platform-Based Digital Services

Saskia Ramm[1](✉), Pierre Grzona[1] and Hagen Jung[2]

[1] Chemnitz University of Technology, 09111 Chemnitz, Germany
{saskia.ramm,pierre.grzona}@mb.tu-chemnitz.de
[2] Institute for Applied Informatics (InfAI) e. V., Goerdelerring 9, 04109 Leipzig, Germany
jung@infai.org

Abstract. While the concept of digital platform (-ecosystems) for the provision of corresponding services has been met with great interest both in the broad re-search community and in practical application, process-based considerations for the selection of digital platform services, which are furthermore supported by artificial intelligence (AI), remain unexplored. However, it is precisely the customer processes in the context of the user experience that play a decisive role in the success of targeted platform solutions. Therefore, this paper describes the development of a method that is specifically focused on the process-based derivation of relevant services for digital, AI-based platforms. To develop our method, we draw on a focus group study that operates in the environment of a current research project for the development of an AI-based networking platform and thus enables a first evaluation of the developed method. With our current results, we are thus not only contributing to the knowledge base around digital platforms and ecosystems in connection with artificial intelligence but are also providing a useful action guide for developing these in practice.

Keywords: Digital Services · Platform Economy · BPMN

1 Introduction

The design of digital services – in the sense of the platform economy of value-added services in addition to pure matching offers for linking customers and service providers on transaction-centred platforms [1] – is subject to a variety of aspects. *Stummeyer* highlights three key aspects that need to be considered during their development. These are, firstly, usability with the focus on the product or the value-added service, secondly, the user experience with the focus on the customer processes and, thirdly, the customer experience with reference to the customer of the respective provider [2]. Essential success factors for the user experience are utility – the subjective benefit for the target group – and usability – the efficient achievement of the goal by the user [2].

C. Zinke-Wehlmann und J. Friedrich (Hrsg.): *First Working Conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow*, AIDRST 2023, Informatik aktuell, S. 51–62, 2024. https://doi.org/10.1007/978-3-658-43705-3_5

Within the research project KISS – AI-based Rapid Supply Network, a data-centric platform is being implemented using artificial intelligence (AI). The platform follows the approach of an AI-based semantic market and networking platform that is intend-ed to provide further value-added services in addition to pure provider-customer networking [3]. The focus of the project is on linking the medical sector with manufacturers from the field of additive manufacturing to increase resilience – especially in times of crisis – through the formation of new value creation networks.

In accordance with the success criteria for the design of digital services mentioned at the beginning, the project aims to examine the extent to which the criteria for user experience could be systematically taken into account. The stakeholders as potential users of the platform, who were identified in the first step by means of a stakeholder analysis, are in the project-specific case employees from the medical-social sector as end customers and institutions in the value chain for additive manufacturing as providers of the platform as well as others. More than 30 different actors and various relevant roles were identified. Each institution itself can be regarded as a separate entity with its own requirements, which needs to be mapped as broadly as possible. In order to map the diverse requirements, an approach was developed within the project to best reflect the potential benefits. As a result, we identified a research question, which is investigated in this paper: *How should an approach be designed to derive the services of an AI networking platform based on processes modelled in Business Process Model and Notation (BPMN)?*

This article, which was developed in the project context based on a focus group study, emphases on the process-based definition of relevant services for AI-based semantic digital platforms, which are gaining more and more interest compared to the development of conventional digital ecosystems. However, how to develop AI-supported and platform-based services has not yet been fully explored in research and development. Challenges are seen in the undeterministic behavior of AI-models and legal issues for example with the copyright for the training data. So, a one-to-one transfer of traditional frameworks seem not feasible. Therefore, we present an iterative approach for this purpose in order to contribute to filling this research gap.

To this end, the remainder of the paper is structured as follows: In the next section, related work is introduced, and important definitions are given, before the research method is described in section three. Section four presents the results of the study and thus the developed approach. Before concluding the article with a summary and the presentation of some ideas for further work, possible application potentials of AI within the framework of the presented approach itself are also discussed.

2 Theoretical Background

2.1 Digital Platform Ecosystems

Digital platforms are generally software-based artefacts that connect different stakeholders and facilitate interaction between them [4]. A basic distinction can be made between two types of digital platforms. On the one hand, there are the digital transaction platforms that primarily facilitate matchmaking and transactions, such as

Amazon or Uber [5]. Providers and consumers can thus use the platform to exchange goods (Amazon) or services (Uber) for money. On the other hand, there are so-called digital innovation platforms that focus on the development of new services by providing interfaces from the platform operator through which extensions to the platform, such as apps (Apple iOS) or browser extensions (Firefox), can be created [4, 5].

These two types are an expression of different perspectives on digital platforms, which are expressed in an economic perspective for transaction platforms and in an engineering or innovative perspective for innovation platforms [6, 7]. Whereas the first focusses on the mediation between two market sides [8], the latter concentrate on the development of the platform to offer innovative new services [5]. In this context, the term platform core is used, which provides basic functionalities such as user management but also the tools that enable extension by third parties (for example APIs or other interfaces) [9]. An extended platform is created in which newly developed modules are provided by so-called complementarians [10].

Abstracted from the perspectives, digital platforms consequently consist of a platform core that provides the basic functionality and becomes an extended platform through complementary services. They are embedded in a (digital) eco-system that includes both platform components and the various stakeholders from the platform owner to the complementor to the customer [10]. The result is that digital platforms bring many actors together and enable an easy exchange of resources (e.g. digital representations of offers) across large distances.

2.2 Methods of Process-Centred Software Development

For the development itself, two main streams can be distinguished: user-centred according to *Draper and Norman* and benefit-centred according to *Constantine and Lockwood*. While the focus of the former approach is on the user and the fulfilment of his or her needs, the latter concentrates on the improvement of the tools to systemically support the tasks to be carried out [11–13].

The aforementioned design paradigms consider the complete scope from the first ideas to the development of the software. The use case described here must be distinguished from this, which combines both models, but also only extends to the step of selecting functional elements.

Considering the development of digital services as a system and software engineering task with a focus on requirements elicitation and management, RE for short, according to ISO/IEC/IEEE 29.148:2018, an iterative three-stage process with subsequent work steps can be assumed [14]:

- Stakeholder Requirements Definition Process
- Requirements Analysis Process
- Architectural Design Process

The first step is to identify the stakeholders with connections to the system over its life cycle, whereby according to the standard, requirements from the management level should also be considered in addition to pure user and developer requirements. The requirements can consist of needs, expectations and wishes. The next stage is the transfer to an implementation-independent model of the future system, which describes it and the degree of fulfilment of the requirements. In the last stage, the system functions would be identified and assigned to the elements of the system architecture, and interfaces and system boundaries would be defined.

The other standards for software development that follow on from this will not be discussed in detail. They describe the relationship between RE, software development and formal vocabulary, but also methods for quality metrics or evaluation of tools for RE [15].

Methods for determining the knowledge of the stakeholders depend on the object of determination but also on the existing sources of requirements. Since this is a new development, document- and system-based methods will not be discussed further. Nevertheless, business process analysis as a document-based method should not remain unmentioned at this point. Suitable investigation techniques according to *Rupp* are, for example, questioning, observation or creativity techniques [16]. In the case of the former, surveys and interviews are the most important. A method for the next step of deriving requirements would be requirements analysis as a multi-stage process as well as testing techniques for the requirements, such as reviews, test cases, prototypes, and analysis models. In parallel, documentation and management should take place [16].

2.3 Documentation of Process Knowledge

In order to record and document both the user and utility requirements resulting from the everyday scope of work, the aim was to be able to do this as intuitively and easily understandable as possible. The focus here was on the defined aspects of utility and usability.

Since the requirements were recorded at the level of business processes and their activities, with a focus on the development of the model, this is the "design model" use case. According to *Van der Aalst*, this has a strong correlation with the choice of process modelling language [17]. The objective of modelling at the conceptual level is linked to essential criteria in the selection decision. These include ease of reading and interpretation, ease of learning and a wide range of application, so that after a comparison based on comparative studies, the decision was made in favor of BPMN 2.0 [18].

3 Methodology

The methodology presented below was developed in an exploratory approach using a focus group consisting of participants who record the needs of the users, process analysts and platform developers. The focus group participants were selected based on their insights and experiences that are highly relevant to the research. Their specific goals were to be incorporated into the method. As a research method, focus groups offer the opportunity to collect data on a specific concept, among other things, as is the case in our study [19, 20]. Due to the different points of view and perspectives of the participants, communication and interaction in an organised discussion provides a flexible means of exploring new ideas or testing new concepts and thus ultimately generating data on the research question [19, 21]. The focus group sessions were conducted according to the procedure of *Bär et al.* in the context of various online appointments [22].

The focus group used in this study came from the KISS project, which aims to develop an approach for AI-supported digital platform services. The research project pursues the approach of enabling the rapid development of value creation networks by means of digital platforms in today's times of dynamic change and thus improving the ability of individual institutions to act. It aims to ultimately increase the resilience of the manufacturing industry through the targeted AI-based networking and to improve the business capability of companies (in the project particularly in additive manufacturing with the goal of scaling to the manufacturing industry) as well as customer sectors affected by significant changes (in the project particularly in the medical and social sectors, possibly expanding to other sectors). However, before a comprehensive technical and organisational concept can be developed for the implementation and successful use of the AI-driven platform, it is first necessary to analyse the relevant ecosystem, considering the value network as a socio-technical system (organisation, processes, data, IT). This results in a large number of relevant stakeholders for the platform, which must be integrated with their specific product, process and service descriptions to form the necessary value chains and analysed with AI support. The Supply Chain Operations Reference Model (SCOR) is a reference for the operation, description and measurement of value creation networks and is therefore used as the basis for the integration mechanisms of the AI-based networking platform to be developed [23]. This is because the digital platform is supposed to generate these value chains and must therefore support the activities that, according to the SCOR model, most companies carry out in order to make their supply chains effective. Based on this, we argue that the process-based view of the different user groups is suitable to identify the services relevant for the digital platform and to provide them with the greatest benefit. Due to the prevailing diversification of the target groups relevant for the platform, a uniform process model must be created that ensures a common procedure for deriving platform-based digital services.

In the following section, we therefore present the method developed in this context by means of a focus group study in a comprehensive way and also go into the individual steps of the procedure model.

4 Approach for the Selection of Relevant Services

Based on the discussions, a multi-stage process was developed in order to specifically focus on the user experience of future users in platform development (**Fig. 1**)

The concept consists of five steps, whereby the first step is repeated iteratively in order to generate a comprehensive database. The individual steps of the concept are explained in more detail below.

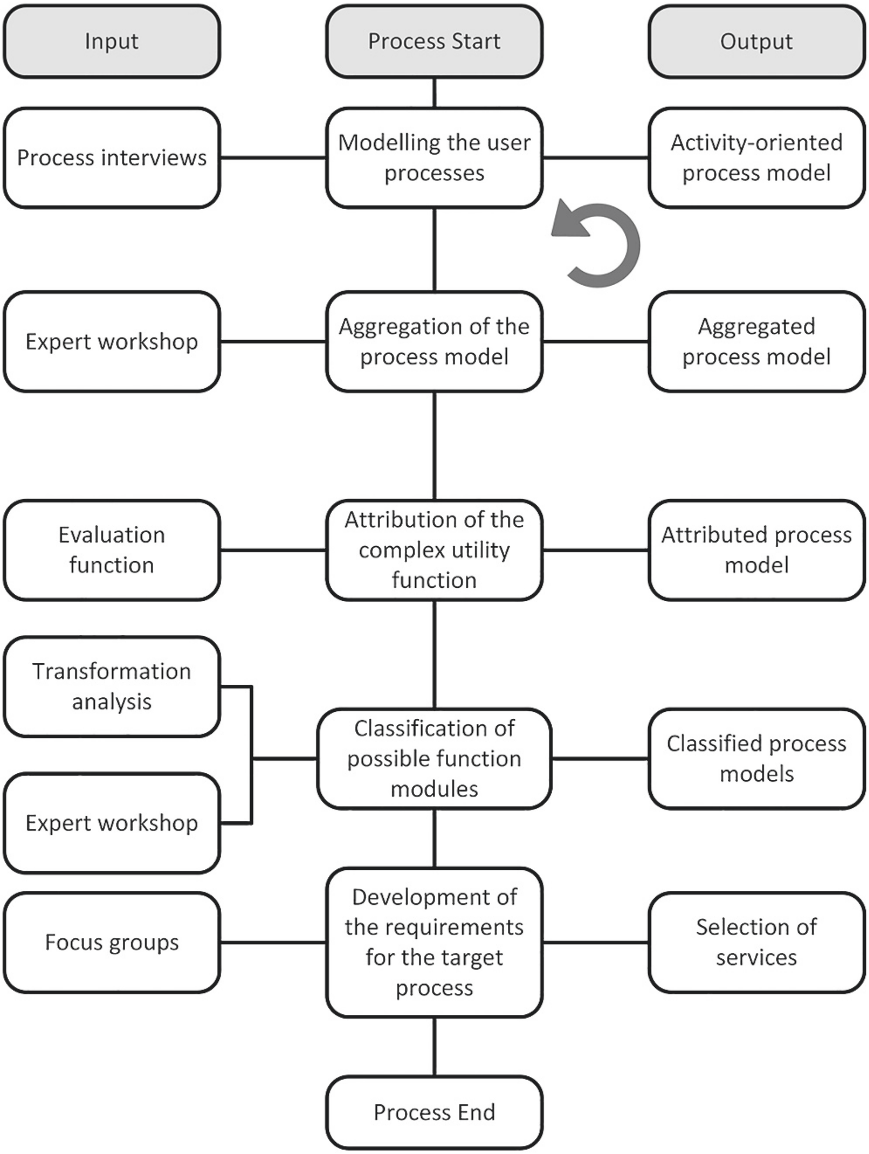

Fig. 1. Concept for the selection of platform-based digital services

4.1 Modelling the User Processes

As already described, the method developed aims to derive meaningful services of a digital platform in a process-based manner. Accordingly, the existing processes of the potential user groups of the platform, which must be adequately recorded and modelled, form the basis of the process model. The aim here is to obtain as detailed a picture as possible of the processes to be modelled in the institutions by means of process interviews. BPMN 2.0 has proven to be a suitable modelling language. These activity-oriented business process models (BPM) with the responsibilities structured in Swimlanes serve as a starting point to document the current effort or the possible need for utility in the respective institution. The process recording in the institutions is conducted in semi-structured interviews, which is a proven method in qualitative research. The questions are not rigidly prescribed, so that the dialogue between the interview partners is as free as possible, which is supported by an interview guideline with prepared questions structured according to topics (e.g., introduction, process initiation, challenges in the process, conclusion). On the one hand, this interview guideline serves as an orientation aid and for structuring the interviews. On the other hand, it also ensures the comparability of the interviews conducted. It is important that the questions are always formulated as openly as possible to stimulate free explanations of the processes and circumstances. Nevertheless, it is recommended to prepare so-called specifications, especially for the very specific questions about the process, which provide approaches for targeted follow-up questions and accordingly help the interviewer to further specify the initial answers. While recording the process, individual and double interviews should preferably be conducted to prevent interviewees from withholding information that they do not share with (larger groups of) third parties, such as superiors or colleagues. This goes hand in hand with the fact that persons with similar responsibilities and positions should always be interviewed across institutions to ensure comparability of process views. Similarly, all interviews should be conducted by the same member of the project team to increase comparability between the interviews conducted. To be able to adequately create the individual BPMs afterwards, we recommend that the interviews are first recorded acoustically and then transcribed.

4.2 Aggregation of Process Models

In the next step, the process models created in the first process step are to be aggregated into a generic BPM in order to combine the various individual processes and thus form activity clusters that are suitable for possible implementation as individual function modules within a platform. By means of expert workshops consisting of process experts and experienced process analysts, a generic BPM can be derived that generalises the individually recorded processes of the individual user groups to simplify the prevailing process diversity. At this point it should be pointed out that this aggregation step is of course accompanied by a certain loss of information from the individual process models of the individual institutions. Nevertheless, this aggregation requires the definition of framework conditions to be observed in order to systematically design this process step. For this purpose, it is conceivable, among other things, to limit the number of activity clusters resulting from the aggregation of the process models in order to reduce the complexity of the aggregated BPM to a reasonable level. Furthermore, we recommend using the lanes as a criterion for the aggregation of activities and accordingly not merging activities across lane and thus role or department boundaries, so that existing process interfaces remain intact. Taking such essential boundary conditions into account, two alternative procedures are available in principle in the course of process aggregation. On the one hand, all captured process models as well as the activities they contain can flow equally into the generic BPM. On the other hand, it is also possible to prioritise the information that enters the process aggregation, for example, to focus on more frequently occurring process paths or to trivialise infrequent activities.

4.3 Attribution of the Complex Utility Function

The previous steps describe an empirical actual state of activities and events of a synthesised existing process. In the following, we want to start from this modelled state and assume that an alternative "better" target process can be found, which is more useful on the one hand and is made possible using digital services on the other. Manual processes should be simplified and substituted by digital sub-processes.

The planned transformations should be cost-benefit efficient. For the improvement of total costs and total benefits, we want to make each activity, each event and each decision assessable in one step. We aim to formulate selective evaluation criteria to estimate the respective benefits at different levels. For example, we would assign attributes such as "temporal -", "monetary -", "crisis resilient –", "emotional benefit" etc. to each activity. In addition to different benefit values of each element of BPM, we also estimate different cost values. Costs can be of various kinds. For target activities to be evaluated, implementation, establishment and operating costs are incurred, among others.

Costs and benefit values do not necessarily have to stand in relation to each other individually. In the next steps, they are used to evaluate potential transformations in terms of whether the overall cost-benefit ratio would increase. To determine this, we propose simulations using a complex benefit function that integrates all cost and benefit attributes in a weighted way. In the simulation, BPM instances are processed, for which additional measured or estimated knowledge about flow distributions is needed.

From a scientific point of view, we find it interesting to investigate the methodology of determining suitable benefit attributes and the joint calculation in a complex benefit function regarding the user experience. The result is an extended and assessable BPM.

4.4 Classification of Possible Function Modules

In the next step, various alternative processes and sub-processes are to be developed and discussed with the help of the previously developed evaluation methods.

New substituted activities proposed by experts need to be estimated in terms of their costs and benefits in the same way as existing activities. It should be noted that the transformation into an improved process is associated with one-off costs. These costs include both the one-off potential development costs of the digital services and the overcoming costs that all the entities involved must incur to accept the change and establish it in the long term. As an example, a potential actor of a platform has to actively decide to give up his usual processes and get motivated to learn new procedures, etc. These costs should be considered in the analysis. If necessary, it may even be necessary to model intermediate transformations, because it is only through these that the new target process appears achievable. All in all, a set of new imaginable target processes should be modelled, which are more useful, more accepted, and not more cost-intensive over a defined period.

Our focus is on categorising all variations of individual functionally definable sub-processes individually and in combination. The resulting set of different potential overall processes can then be ranked using the proposed evaluation. We finally select the best-rated, most useful, and feasible alternative process including the new sub-processes.

4.5 Development of the Requirements for the Target Process

The last step in the process model describes the derivation of the requirements of the target process in the context of a focus group. According to *Rupp*, this can also be seen as a review of the requirements [16]. This workshop takes place with both developers and stakeholders – the future users. The users should have the same role and sector affiliation as in the process interviews, since they were identified as relevant stakeholders at the beginning. In the focus groups, it will be examined whether the classified process works with the future role distributions on the one hand, and on the other hand, which building blocks could be considered for a possible implementation. This functional selection of possible services could then be the starting point for further development and keep the effort low for all those involved.

5 Conclusion and Outlook

In general, digital platforms and the ecosystems that are inevitably linked to them are not only part of a wide variety of research disciplines, but they have also aroused great interest in the past, especially in business practice. Nevertheless, the systematic derivation of beneficial services for digital platforms, especially in connection with the use of AI, remains undeveloped. We therefore propose the idea of a process-based approach to deriving platform-based digital services, which is reported in this paper. In this context, the findings from this focus group study are not only of interest to design scholars (the authors or others) who are interested in developing such a platform and thus in designing the method. They equally contribute to pointing out a development and implementation agenda for future users of such AI-based platforms and their developers. Focusing on utility and usability with an emphasis on subjective utility and efficient goal attainment can effectively narrow the solution space to be considered as an aid in this regard.

However, the process model just presented not only provides the basis for deriving suitable digital services for an AI-based platform on the basis of real processes. On the other hand, it also itself presents various potentials for the use of AI to support the individual steps of the method, which are currently envisaged within the framework of empirical methods and would consequently experience a significant reduction in workload through applications of AI. For example, the second step of the procedure, in which the aggregation of the individual process models (see 4.2) of the institutions under consideration takes place, opens up the possibility of carrying this out independently, e.g. rule-based, analogous to the SCOR model. Likewise, the fourth step in the classification of possible function modules (see 4.4) shows another possible use of AI solutions in the context of a cluster analysis. This already shows the first starting points for the further development of the method presented. Although the process model is undergoing initial validation as part of the KISS project, the design process has not been fully completed and thus creates the starting point for further research.

Our research focus in this work is on the elaboration of digital processes as part of the user experience rather than the design of the user interfaces. It should also be mentioned that the respective framework conditions are a considerable limiting factor. On the one hand, only a predefined section is surveyed during the process recording in interviews and thus all further results are pre-determined to a certain extent, and on the other hand, the number of decision-making options within each step is limited by the framework conditions, so that it may well be that no functional changes are considered beneficial. The composition of the respective focus and expert groups also determines success.

Acknowledgment. The Project "KISS – AI-based Rapid Supply Network" (FKZ: 01MK22001B) is supported by the Federal Ministry for Economic Affairs and Climate Action (BMWK) on the basis of a decision by the German Bundestag.

References

1. von Engelhardt, S., Wangler, L., & Wischmann, S. (2017). *Eigenschaften und Erfolgsfaktoren digitaler Plattformen. Eine Studie im Rahmen der Begleitforschung zum Technologieprogramm AUTONOMIK für Industrie 4.0 des Bundesministeriums für Wirtschaft und Energie*.
2. Stummeyer, C. (2022). *Ausgewählte Aspekte bei der Gestaltung von Digital Services. In: Digitalisierung in Industrie-, Handels- und Dienstleistungsunternehmen* (S. 77–89). Springer Gabler.
3. BMWK. (2023). KISS – KI-gestütztes Rapid Supply Network. KI-basierte semantische Plattform zur rapiden Etablierung von Wertschöpfungsnetzwerken in Krisenzeiten. https://www.digitale-technolo-gien.de/DT/Redaktion/DE/Standardartikel/KuenstlicheIntelligenz-Projekte/KuenstlicheIntelligenz_Alle_Projekte/ki-projekt_KISS.html. Zugegriffen: 9. Febr. 2023.
4. Tiwana, A., Konsynski, B., & Bush, A. A. (2010). Research commentary —Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information Systems Research*. https://doi.org/10.1287/isre.1100.0323.
5. Bonina, C., Koskinen, K., Eaton, B., & Gawer, A. (2021). Digital platforms for development: Foundations and research agenda. *Information Systems Journal*. https://doi.org/10.1111/isj.12326.
6. de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: A research agenda. *Journal of Information Technology*. https://doi.org/10.1057/s41265-016-0033-3.

7. Rolland, K. H., Mathiassen, L., & Rai, A. (2018). Managing digital platforms in user organizations: The interactions between digital options and digital debt. *Information Systems Research*. https://doi.org/10.1287/isre.2018.0788.
8. Malgonde, O., Zhang, H., Padmanabhan, B., & Limayem, M. (2020). Taming the complexity in search matching: Two-sided recommender systems on digital plat-forms. MISQ. https://doi.org/10.25300/MISQ/2020/14424..
9. Spagnoletti, P., Resca, A., & Lee, G. (2015). A design theory for digital platforms supporting online communities: A multiple case study. *Journal of Information Technology*. https://doi.org/10.1057/jit.2014.37.
10. Hein, A., Schreieck, M., Riasanow, T., Setzke, D. S., Wiesche, M., Böhm, M., & Krcmar, H. (2020). Digital platform ecosystems. *Electron Markets*. https://doi.org/10.1007/s12525-019-00377-4.
11. Constantine, L. L., & Lockwood, L. (2002). Usage-centered engineering for web applications. *IEEE Software*. https://doi.org/10.1109/52.991331.
12. Constantine, L. L., & Lockwood, L. A. D. (1999). Software for use: A practical guide to the models and methods of usage-centered design. Pearson Education.
13. Draper, S. W., & Norman, D. A. (1986). *User centered system design: new perspectives on human-computer interaction*. L. Erlbaum Associates.
14. ISO International Organization for Standardization. (2018). Systems and software engineering – Life cycle processes – Requirements engineering **35.080**(ISO/IEC/IEEE 29148).
15. Schneider, F., & Berenbach, B. (2013). A literature survey on international standards for systems requirements engineering. *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2013.01.083.
16. Rupp, C. (2020). *Requirements-engineering und -management*. Carl Hanser Verlag GmbH & Co. KG.
17. van der Aalst, W. M. P. (2013). Business process management: A comprehensive survey. *ISRN Software Engineering*. https://doi.org/10.1155/2013/507984.
18. Pereira, J. L., & Silva, D. (2016). Business process modeling languages: A comparative framework. *New Advances in Information Systems and Technologies* (S. 619–628). Cham: Springer.
19. Stewart, D.W., & Shamdasani, P.N. (2015). Focus groups. Theory and practice, 3rd edn. Applied Social Research Methods Series, vol. 20. Sage, Los Angeles (California) [etc.].
20. Williams, M. (2003). Making sense of social research. SAGE Publications.
21. Fern, E. F. (1982). The use of focus groups for idea generation: The effects of group size, acquaintanceship, and moderator on response quantity and quality. *Journal of Marketing Research*. https://doi.org/10.2307/3151525.
22. Bär, G., Kasberg, A., Geers, S., & Clar, C. (2020). Fokusgruppen in der partizipativen Forschung. *Partizipative Forschung* (S. 207–232). Springer VS.
23. APICS. (2017). Supply Chain Operations Reference Model SCOR: Version 12.0.

Opportunities of AI for society

Abstract. The opportunities of technological innovation and their impact on the society will be highlighted in the contributions of this session. The presented approaches show how AI-based technologies can contribute to solving societal challenges, e.g. in law enforcement. Ethical aspects will be discussed and questions about responsibility and transparency in dealing with AI will be raised. The session will provide space for a broad dialogue on the impact of AI on society and possible strategies to make the most of the benefits of AI.

Classification of Static Poses Based on Key Point Detection for Application of Incriminated Image Files

Schönbrodt Antonia[1,2](✉)

[1] DigiFors GmbH, Torgauerstraße 231, 04347 Leipzig, Germany
[2] Springer Heidelberg, Tiergartenstr. 17, 69121 Heidelberg, Germany
info@digifors.de
https://digifors.de/kontakt/,
http://www.springer.com/gp/computer-science/lncs

Abstract. The role of artificial intelligence, particularly in enhancing decision-making processes and facilitating automation, has become indispensable in today's society. Law enforcement agencies have used the findings of this research for several years to analyze the ever-growing volumes of data. This approach owns an important role in the detection of child and adolescent pornography. Research conducted in this field has traditionally been dependent on the findings from skin filtering studies. However, some scientific publications and empirical studies reveal insufficient classification when using nudity levels. This work takes a more promising approach by leveraging the knowledge of an existing system and combining it with a homegrown model to improve the detection of incriminated image files. For this implementation, techniques from motion analysis were used to open up a new field of pose recognition.

Zusammenfassung. Die Rolle der künstlichen Intelligenz in Bezug auf ihre Individualität und Automatisierung ist in der heutigen Gesellschaft unverzichtbar geworden. Die Ergebnisse dieser Forschung werden seit einigen Jahren ebenfalls genutzt, um Strafverfolgungsbehörden bei der Analyse der ständig wachsenden Datenmengen zu unterstützen. Dieser Ansatz besitzt eine wichtige Rolle bei der Aufdeckung von Kinder- und Jugendpornografie. Bisherige Studien in diesem Bereich stützen sich größtenteils auf die Ergebnisse von Studien zur Hautfilterung. Einige wissenschaftliche Veröffentlichungen und empirische Studien zeigen jedoch, dass die Verwendung des Nacktheitsgrades keine ausreichende Klassifizierung ermög-licht. Diese Arbeit verfolgt einen vielversprechenderen Ansatz, indem sie das Wissen eines bestehenden Systems nutzt und es mit einem selbst entwickelten Modell kombiniert, um die Erkennung

Supported by DigiFors GmbH.

C. Zinke-Wehlmann und J. Friedrich (Hrsg.), *First Working Conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow*, AIDRST 2023, Informatik aktuell, S. 65–82, https://doi.org/10.1007/978-3-658-43705-3_6

inkriminierter Bilddateien zu verbessern. Für diese Implementierung wurden Techniken aus der Bewegungsanalyse verwendet, um ein neues Feld der Posenerkennung zu erschließen.

Schlüsselwörter Artificial intelligence · image categorization · pose detection · data pre-processing

1 Introduction

The idea of using artificial intelligence to assist humans and introduce new innovations has been around since its inception in 1950 [1]. Its use has already consolidated and is constantly expanding in many fields such as business, the automotive industry, and many others [2]. Nevertheless, the field of artificial intelligence is not sufficiently considered in the context of law enforcement. As digitization continues, ever-increasing amounts of data are being generated that cannot be handled by law enforcement's own resources alone [3]. In order to facilitate the analysis process for investigators, a field opened up in academia a few years ago in which neural decision patterns are used to examine and evaluate the volumes of data. The surge in data volume is especially noteworthy in less-visible crime areas. This is also true for the analysis and detection of child and adolescent pornography [4]. Here, not only can the analytical aspects be made more efficient through the use of neural networks, but artificial intelligence is also used to minimize the psychological burden on investigators. However, most of the models to be tested are currently still in the testing phase. For this reason, a bachelor's thesis was written to provide a foundation for the research listed here. It is intended to change the perspective regarding the use of artificial intelligence for the detection of child and adolescent pornography. To this end, existing methods have been partially redesigned. This approach makes them more resilient to large amounts of data, as this adapts the methods to other subjects. In this way, the constructed systems can be adapted to the different needs and interests of the users in order to revolutionize the analysis of the addressed offense domain. Thus, the goal of this thesis is to address the question of whether criminal image recognition can be improved by allowing knowledge of pose recognition to have a decisive influence on the determination of incriminated images. The basis for this theory is the aforementioned bachelor thesis, which was further adapted and improved after its completion.

2 Scientific Framework

2.1 State of Research – Artificial Neural Networks for Child Pornography Detection

The emerging field of artificial intelligence in science now makes it possible to analyze large amounts of data faster and more efficiently and in a more targeted manner[5]. The most promising scientific contributions are currently still in the research stage [6]. In preparation for the aforementioned bachelor's thesis, systems dealing with this area of crime were examined in a practical manner in police agencies. None of the models tested were targeted, as most research in this area relied on the use of the skin filter alone. A suitable scientific publication in this regard was written by Nicole Garbers and Michael Brodthage[6]. This paper delves into the challenges posed by employing artificial intelligence in the detection of child pornography content. In their 2021 execution, they describe the limitations of the skin filter often used for classification purposes and their improvements. According to them, the skin filter by itself does not provide suitable preselection [6]. One problem that arises from the nudity level classification is that the neural networks are not able to recognize different ethnicities [7]. On the other hand, misinterpretations of the analyzed data also occur in everyday situations of the data under analysis. This phenomenon of misclassification also occurs when selfies are taken. Since these image recordings represent a high percentage of skin of the entire image [8]. In addition, borderline cases often occur on the evidence being analyzed, where underage individuals are required to pose sexually clothed. These considerations show that the skin filter determinations for pornography detection [9] cannot be used alone in such cases. Moreover, the described views of Garbers & Brodthage could be confirmed by the results of the apparent observations of the practical models. The multilayeredness and complexity of the phenomenon area to be addressed and the listed limitations of Gabers and Brothage have the consequence that the analysis of the nudity level has to be improved by considering further characteristics. Since final thesis only provide a limited framework for research, the considerations in this area must be narrowed down. The scientifically most important area is provided by the motion analysis with the analysis of the joint points [10], [11], [12]. Despite the scientific dissemination of this topic, the determination of poses and their classification was not recognized in court for a long time. This changed with the 49th Criminal Law Amendment Act [13]. Through this amendment, recordings without a direct sexual act are now considered criminal. It is now sufficient only to detect the pose itself [13]. Based on the arguments described above, the question of whether the detection of criminally relevant pornographic material can be improved by determining and classifying static poses will be explored.

2.2 Pose Detection

According to the Criminal Code [§184 para. 1 no. 1 lit. B], posing of a partially or fully clothed child is punishable. In this context, the legislator defines posing as an "unnatural and sexually accentuated posture". In this category, children are induced to assume sex-related and stimulating postures in order to imitate adult models [13]. For the detection of posing, a uniform determination scheme must be established first. In the context of this work, unnatural poses are static postures that do not correspond to the typical behavioral patterns of children. This category includes positioning the buttocks in front of the camera, rolling the body on the floor and spreading the legs. In addition, there are targeted postures in which certain body regions are deliberately accentuated by certain inclinations of the arms or legs. Furthermore, images in which touching the chest or intimate area is in the foreground are counted under a gender-emphasized posture. According to [14], tracking and estimating the human pose is a computer vision task. The basis of this is the recognition, assignment and tracking of key semantic points. There are already exist various software applications to support the fixing of joint points and thus the determination of poses. During the research period, different applications were considered and evaluated. Various pose detection systems were tested for the planned procedure and were adapted to the conditions. The major known pose detection systems AlphaPose [15] and Mask R-CNN [16] are not suitable for the intended classification. Mask R-CNN is based on semantic segmentation, which can lead to shallow human bounding boxes and this leads to a incorrect classification of joints. Furthermore the system was fundamentally not designed for pose estimation [17]. The AlphaPose pose estimation system is rather designed for the determination of multiple individuals. The scientific review by Siddharth Sharma also mentions that this model provides more failure cases than other models in pose estimation [17]. In order to actively compare the favored pose estimation system OpenPose with another suitable system, local tests between OpenPose and another system called MediaPipe were performed as part of the bachelor thesis developed. It has been shown that OpenPose can detect the skeletons more reliably. In particular, hidden body parts could be detected better as with MediaPipe due to the additional function of occlusion.

2.3 OpenPose

OpenPose is a pose recognition system [18] which was developed based on the COCO dataset. OpenPose is the first "open source real-time system for 2D pose estimation" [18], which is able to determine key points of the human body. The keypoints describe various points that are placed at significant locations on the body, such as the shoulders, knees, and elbows. [19] The basic principle of the convolutional neuron network based approach is the so-called bottom-up approach. Here, in a given image, the different keypoints are first determined and then the differently recognized skeletal parts are combined to form a whole object [19]. Such approaches exhibit a certain robustness and have the potential to decouple the influence of runtime complexity from the number of people [20]. OpenPose provides the first bottom-up approach in which association values are used to encode an image in terms of limb position and orientation via so-called Part Affinity Fields (Pafs) [18]. Here, the Pafs are defined as a set of 2D vector fields consisting of a series of flow fields. [18]. After the initial estimation of these vectors, the information is improved based on the multilevel classifier at each refinement level. Subsequently, a grouping procedure is performed in which the vectors corresponding to the same joint region are connected and ranked according to affinity. This principle is based on the idea that the joints to be detected are connected by limbs. Finally, the detected points are connected to represent the pose key points. First, the connections of one group of joints are performed and then those of the other limbs are determined [18] Thus, the connection of the body-and-foot keypoints detector finally provides a complete human skeleton with at least 25 different keypoints, which can be used for further considerations.

3 Methods – General Procedure

3.1 Assembling the Data Material

Different data sets were compiled for pose recognition and neural network training purposes. Given legal and ethical constraints, we cannot use incriminated image files for creating suitable training material, as the possession of child and youth pornographic content remains prohibited, even for scientific endeavors [12]. For the work with artificial neural networks, various data sets are needed, which due to their diversity ensure a robustness of the network against large amounts of data. Initially, we require a dataset tasked with training and updating the neural network. In order to achieve diversification of the dataset, care must be taken to use shots of people of different ethnicities, group shots, and backgrounds in the classes under consideration [6]. In addition, diverse shooting angles and situations have to be considered in the image scenes to under consideration.The freely available collection of Figshare [21] was used as the pornographic data set. However, due to this, the number of usable image files reduced rappide. Since the goal of this research is to distinguish between posing and no posing images, everyday images were added to the dataset as an opposing class. Thus, both domains form the training dataset used in this work. Considering the limited time for the underlying bachelor thesis and the limited datasets, only the small amount of data was relied on. In order to check how well the learned model generalizes the data or if it represents only the training data, a validation dataset is necessary. For this purpose, the corresponding data from Figshare [21] was also taken into consideration. Then, another independent dataset is needed to test the neural network. The files from the test dataset of Figshare [21] are not only pornographic images, but also everyday clippings. The use of both classes is necessary to test the functionality of the artificial intelligence related to the objective. Only in the comparison between non-pornographic and pornographic poses it can be seen if the system performs a good classification.

3.2 Created Workflow

The scientific thesis of the underlying bachelor thesis is to be able to make a classification of sexual poses on the basis of defined cornerstones. For this purpose, a workflow adapted to this topic was developed, which already uses an existing AI-generated innovation and combines the basic knowledge with a self-created instance.

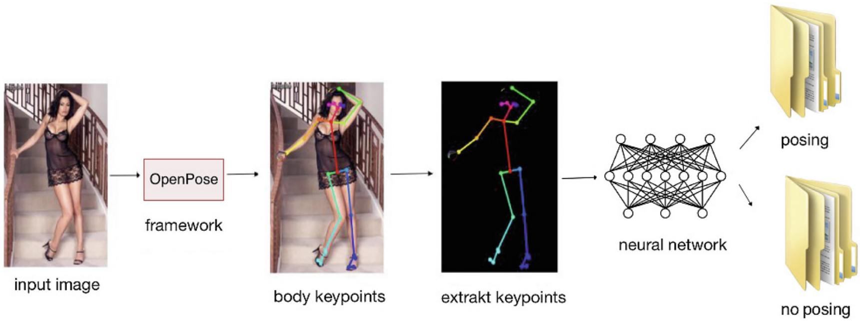

Fig. 1. Visualization of the created pipeline with included integration of the two neural systems

This compiled construct provides information about whether a human pose is understood as posing or not posing. The listed parts were combined in a Python code to form a packaged analysis system. The architecture of the workflow is shown in Fig. 1. At the beginning, the pipeline is briefly summarized for a better understanding. First, the image files to be analyzed are passed to the program. In this process, the files are first examined using the embedded framework OpenPose, where the person-based rigs are created. The results are optimized with respect to the objective. Subsequently, they are passed to a neural network, which was constructed self-referentially for the problem to be analyzed. The structure of this system is described in more detail below. This network divides the image data into the classes *posing* and *no posing* and then assigns them to equal folders, which can be viewed at the end of the pipeline with the respective results. The entire process was visually structured so that at the end it is comprehensible which image files were classified as posing and which need to be examined more closely by the officials. The output of the workflow is therefore not the human skeletons created in the intermediate steps, but the data originally entered. This form of representation ensures that each image can be uniquely assigned. Thus, the decision of the neural network can be better observed and evaluated. Possible misinterpretations in the representation of the original files can be better detected. Thereupon, a faster identification and adjustment of the inaccuracies can take place. These deviating results are then used to be integrated into the new learning process of the neural network and to improve it.

Prediction of person-specific body skeletons by OpenPose. The OpenPose motion analysis system was used for pose determination. In this study, only the demo version of this implementation was used, as it does not rely on an NVIDA graphics card, unlike the more comprehensive version. The initial testing did not yet yield the possibility to test the encryption results with the full version. Whether changing the way the program is run improves the results of this work will be tested in further future research. Through the demo version used, human skeletons, also called rigs, are generated from the incriminated data. To better narrow the predefined skeletons for the range of phenomena to be considered, slight changes were made to the implementation code line of OpenPose.

Listing 6.1. Representation of the customized demo code line from OpenPose

```
bin\OpenPoseDemo.exe --image_dir examples/media/
--disable_blending --hand --write_image folder_path
```

The parameters were changed by the attribute *disable_blending* to remove the background from the images to be analyzed. This change allows the neural network to analyze only the keypoints and avoids distortions caused by accentuated backgrounds. In addition, the number of keypoints to be classified has been increased by also inducing the detection of the feet and hands. The foot detection is already integrated by default in the demo version. With the attribute *–hand* the hand detection was added to the determination.These body regions play an important role in the detection of posing images. Since especially hands can provide clues for a conscious accentuated movement. However, not all poses are uniquely recognized when OpenPose detects the key points. For example, there are often deviations when the spine is bent, the legs are stretched in the air, or the buttocks are held up to the camera. Currently, attempts are being made to counteract this problem by specifically training the neural network with such borderline cases in order to promote the robustness of the network. On the other hand, efforts are also underway to develop the Linux-based variant of the workflow to incorporate the deeper innovation of OpenPose.

Labeling of the inserted data sets – preparation of the training phase. To initiate the determination of the poses, they must first be precisely evaluated and categorized for training purposes. These differences are then learned by the neural network so that it can determine a suitable classification for unknown image objects. Reliable preparations are pivotal for the success of training. After the initial billfiles were pipelined, OpenPose added keypoints to them, and due to the change in code implementation, the background was subsequently removed. All that is left is the pure skeleton. Subsequently, in preparation for the training, the values of individual skeletons must be established. For the labeling of the training data exclusively results are used, with which OpenPose could detect nearly all key points. Such well-detected skeletons include outputs in which there are hardly any deviations from the original keyponits detection. The absence of single small but still visibly complementary body parts, such as the absence of a lower leg, is within the tolerance range. A visualization of such good results is shown in Fig. 2. Based on this selection a robustness of the later trained network is generated and misinterpretations are prevented.

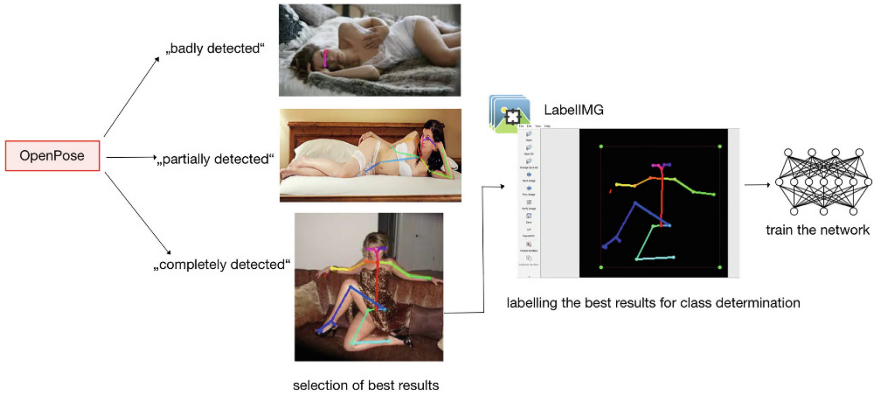

Fig. 2. Visualization of labeling approach in preparation for neural network training

For the assignment of the two classes *posing* and *no posing* the method of labeling is used. In this method, characteristic sexual poses are selected and assigned to the class *posing*. For the classification of data, different applications exist to determine the respective images or to assign them to different classes. For this work the tool LabelIMG was used. With this tool the image files with the most recognized keypoints are read in and assigned to the desired class by hand. The file is provided with a bounding box, similar to a bounding box. This frame is then assigned the class *posing* or *no posing*. This assignment works similar to the assignment of object views, where objects such as ships are assigned to the class *ship* due to their nature. The class information and coordinates of the boxes are saved and stored in an xml file. These for the further procedure used and afterwards the own developed net passed.

Construction and training of the neural network As a proof of concept for classification, a simple neural network with a convolutional layer based on the existing OpenPose system was attached to process the results of the herding system. The results of this network, however, did not provide satisfactory results, because due to the architecture and the resulting output only one class was detected. As a result, the proof of concept was adapted, revised and transformed into a convolutional neural network with pooling layers during the course of the research. The individual layers of the self-developed transformed network are shown in Fig. 3. To increase the processing power of the system, the output image files of the previous mesh were initially scaled to an input value for the new mesh of 64 x 64 pixels using the *Resize (64,64)* function. This resolution was chosen to increase the processing power of the mesh.

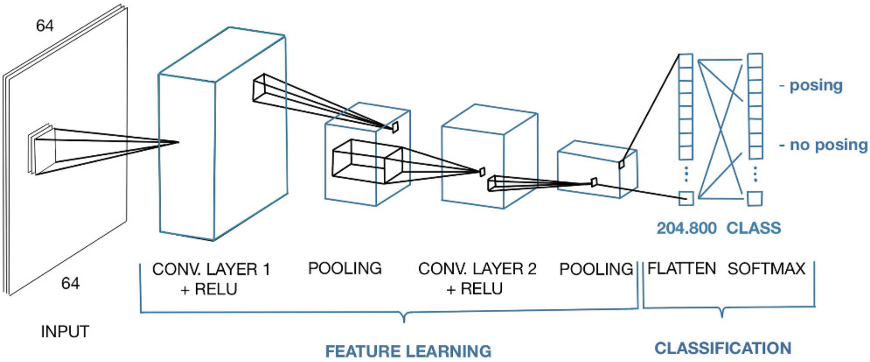

Fig. 3. Representation of the individual layers of the specially created neural network

Previous studies frequently utilized higher resolutions as they facilitate the extraction of more information from an image. Nonetheless, as this study is primarily focused on pose detection rather than detailed image analysis, a lower resolution was selected. In addition, the input layer is represented by 3 layers, for the image colors red, green and blue. These pixels are responsible for making colors visible on the input files. After the input layer, two Convolutional Layers are added. Their output is determined by the so-called ReLu function. They follow the principle of feature learning. Subsequently, the data is reduced by the adjacent pooling layer. This is done by filtering out the strong features and discarding the weak features. At the end of the network there are two linear layers, which are transformed into a one-dimensional layer with the help of the *view()* function. These layers are finally responsible for the classification. After defining the neural network structure, the forward method is applied in the program code to connect the network architecture. It also enables the execution of the forward pass. Consequently, the layers process together until the output is available, which provides the image. In the output layer, the softmax function is used as the activation function instead of the previously used ReLU function. In this research, the softmax function was deliberately chosen because it is specifically used for classification problems. Many multilayer neural networks end up with real-valued results output at the last hidden layer. These often pose scaling challenges for the neural network, creating obstacles for further processing. Here the softmax function is helpful, since it converts the results into a probability distribution, which can be better used by the network. For the analysis process, the zero points of the gradient and the activation of the weights are determined. Using Pytorch's Autograd module, differentiations are performed and weights of the model are automatically updated. In order for Pytorch to not only determine whether the network's predictions are true or false, but also to determine their valence, a loss function is needed. For this study, we utilized the CrossEntropy function. This is also used in other work for classification tasks. The loss function is used during the training loop to match the predictions with the actual labels. This information is backwarded by the *backward()* method to allow the initial layers to learn from the differences and update the network. Through the *optimizer() function*, the differences are passed to the gradient, resulting in an adjustment of the weights at each layer. For the training loop, the parameters just described are passed to the network, which is shown in the underlying representation. In the code sequences, *Cnnet* was used to describe the self-created neural network.

Listing 6.2. Introduction of loss function and optimizer into training

```
optimizer = optim.Adam(cnnnet.parameters(), lr=INIT_LR)
loss_fn = nn.CrossEntropyLoss()
```

Due to the given resources and the limited time for this research, an epoch of 50 was decided upon. This implies that the training process iterates 50 times. With this number, representative results are already produced, which, however, do not yet drastically extend the analysis process. After all, the accuracy of the training as well as a loss function were determined in order to identify and subsequently evaluate the functionality. In summary, the training can be described as the following lines of code.

Listing 6.3. Representation of the training loop

```
for data, target in tqdm(train_data_loader, desc="Training..."):
optimizer.zero_grad()
output = cnnnet(data)
loss = loss_fn(output.squeeze(), target.float())
loss.backward()
optimizer.step()
#calculate the training loss
H["train_loss"].append(loss.item())
train_correct += (output.argmax(1) == target.argmax(1)).type(torch.float).sum().item()
#calculate the accuracy after training loop
_train_correct = train_correct / len(train_data_loader.dataset)
H["train_acc"].append(_train_correct * 100)
```

However, if only training data is used for programming, the neural network used may overfit. This means that the system is good at recognizing and evaluating representations that it has been trained with. However, generalized representations that are not included in the training dataset cannot be analyzed. For this reason, other datasets such as the validation dataset and the test dataset are mandatory. To fulfill this scenario, a validation loop is created after the training loop. This can ensure that the model is not over-fitted to the training data and recognizes not only them, but also independent data. The training procedure just described is repeated with the validation data. Once again, the dataset undergoes batch-by-batch processing within the system, with the loss being calculated. However, no *backward()* method and optimization approaches are needed here, since the parameters of the model do not change in this step. A similar procedure is also used to pass the data to the test loop. Evaluation is performed independently of training or validation. In order to determine the learning and recognition success of the respective phases, the accuracy of the respective predictions were measured for the training and validation data set and were output after the respective phases. Thus, after each training phase, a percentage is shown which indicates how well the system classified the data. Afterwards, the trained model can be passed to the pipeline.

Composition of the individual components After training the self-generated neural network, the two systems must be connected to form a single unit in order to pass the results of the motion analysis system to the classifier. For this purpose, the two sequences must be merged in the code pipeline. In order to be able to guarantee further modifications of the created network, the architecture and the training of the network were deliberately combined into a separate program code, which was subsequently saved in a file named *new_model.pt*. This project file is then passed to the pipeline in only one program line. This ensures that other possibly better meshes can be trained separately and can later be integrated into the pipeline without much effort. Thus, the pipeline can be quickly and efficiently replaced by even better neural networks, increasing its individuality. Based on the training, the neural network can now evaluate the captured image files and assign them to the two classes *Posing* and *no Posing* to be determined. This assignment is done by sorting the images of the respective classes into two folders of the same name. Since the output of the neural network is the individual skeletons without background, these files should actually be in the folders of the same name. According to the workflow, these folders now contain the images in the form of skeletons. However, if you are not intensively involved with the subject, you will not be able to do anything with this kind of representation. In the meantime, a program has been created that ensures that the converted files are back in the folders as original files at the end of viewing. However, OpenPose automatically changes the file extension of the images as soon as the skeleton is determined. Thus, standard extensions like .jpg, .png or .bmp are replaced by _rendered.png. This procedure makes it difficult to reconstruct the original files. In the underlying bachelor thesis, an attempt was made to solve the problem by removing the file extension _rendered.png and replacing it with some kind of tariff system with .png and .jpg. This led to the fact that 34% of the image files to be analyzed could not be determined. This is because these were not taken into account when restoring the image extensions. To counteract this problem, in later studies a working directory was created containing a folder in which all image files were stored again before the corner points were determined. In this process, all image files were given a duplicate file extension. This means that the file neuronalesNetz.webp became neuronalesNetz.webp.webp after this step. This newly named file is now packed into an intermediate folder *"Zwischenstand"*. Afterwards, the data is passed to the OpenPose system for item recognition. The file extension is now replaced by the predefined phrase again. Thus, the file name is now neuralNet.webp _rendered.png. With this in mind, a Python program was written with the Pose Estimation System extension removed. What remains is the extension of the original file. So the ending is known and at the end all files can be sorted into the respective label folders again.

4 Results

The results of the neural network training provided satisfactory results. At the beginning of the training, the accuracy of the predictions was only 50.75% for the training data set and 52.64% for the validation data set. These values could be improved strongly after the 50 training runs. It was found that the model predicted the values of the training data set with an accuracy of 97.59%. The accuracy for the fit to the validation data set was 98.51%. The loss functions also obtained improved from 0.687834 to 0.085758 for the training data set and from 0.702857 to 0.024533 for the validation data set. The total training time was 3425.92 seconds. By restructuring and adjusting the pipeline, the developed workflow could be improved. For testing purposes, the corresponding reduced dataset from Figshare [21] with 532 data was introduced into the created pipeline.

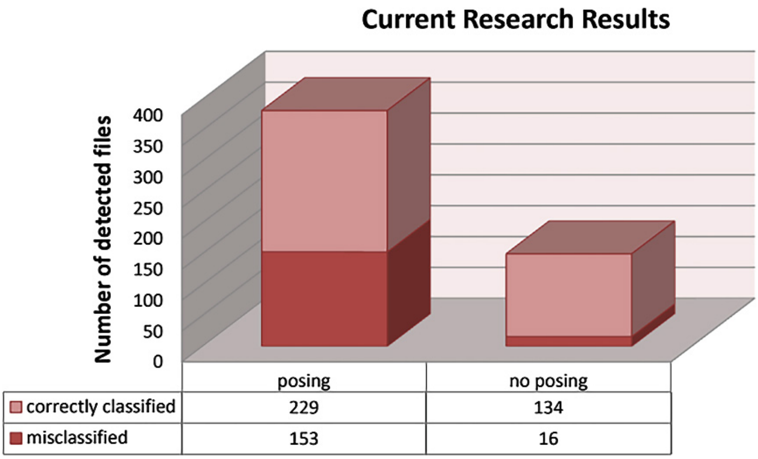

Fig. 4. Graphical representation obtained of the classification results

Due to the change to determine the file extensions, it was now possible to recognize and subsequently determine all image files from the constructed model, unlike in the underlying bachelor thesis. In total, 150 images were placed in the *no* *posing* folder and 382 images were placed in the *posing* section. Fig. 4 illustrates the exact results. It was found that the basic idea of this research, to detect posing on the basis of pose estimation systems, could be applied to many images. In addition, the consideration of the revised output representation could be fully implemented. It was shown that the developed pipeline could reliably classify posing images as such even if the depicted persons showed postures where, for example, the legs are spread or the buttocks are held into the camera. However, the developed system faced challenges in recognizing postures involving bent legs and a curved spine This finding is reflected in the 16 misclassified image files in the *no posing* category. The image files there predominantly illustrate the problem just described. In order to minimize this already known problem, the neural network is currently being trained with such problematic image files in order to improve the recognition rate of these poses. On the other hand, it has been shown that the trained neural network exhibits a slight over-fitting. This means that many apparently everyday poses, 153 in total, were sorted into the category *posing*. In most of the images, the people are working with their hands to express certain situations. Due to the targeted training on the hand region to detect a playful accentuation of the hands, certain hand positions were misinterpreted by the system. In addition, the studies conducted showed that the OpenPose system, which was integrated first, had problems identifying the corresponding skeletons for persons without a recognizable eye region or with the half of the body covered. For example, no skeletons existed for individuals with no eyes visible. As a consequence, the subsequently attached neural network fails to detect potentially sexually suggestive poses as *posing*, if Open Pose doesn't provide a substantial skeleton beforehand After the results were obtained for applied incriminated dataset with adult subjects, a dataset with everyday image files of children was injected into the pipeline. This step was performed in order to verify whether the transfer, of the already confirmed functioning of the pipeline, is also transferable to the proportions of children. Particular attention was paid to the fact that the interpretation of skeletal parts was transferred. The tests showed that both systems implement the learned factors and adapt them to the respective image files.

5 Discussion and Limitations

The results of the research have shown that motion analysis methods can also be used for the detection of sexually suggestive poses. Furthermore, the generated pipeline proved that the estimation system just described can be combined with another neural network. Since the present work deals exclusively with the analysis of movements with a pornographic context, a fundamental improvement in the recognition of child and adolescent pornographic files cannot be confirmed by this one method alone. However, to further fill the gap in scholarship in this area, the proposed methodology must be combined with the other preliminary considerations from the research. This work will pave the way to bring more attention to the issue of child and adolescent pornography and provide further suggestions for the outdated rating systems.

5.1 Classification Problem

An existing problem, which has been shown in the course of this research and on the basis of the results, is that the self-created model can only act in combination with previously switched network OpenPose. This means that if OpenPose provides weak or no keypoints, the subsequent model cannot perform reliable classification either. Although the supposed best evaluation system was chosen for this research, it also has its limitations. The system was developed to recognize human poses and make predictions. However, there are no scientific publications yet on whether this system is also used to detect sexual posing. Thus, the present results are the first attempt to implement the prediction of skeletons by OpenPose in a different domain. It could be confirmed that OpenPose shows problems when the spine of the respective person is curved. Thus, the results were only inconsistent for certain lying positions. In addition, images where no eyes are visible cannot be skeletonized. This means that close-ups of buttocks or genitals are not analyzed when posing. These inconsistencies might have arisen because the actual functionality of the rating system is to recognize people in everyday situations. Should OpenPose become further established for the analysis of pornographic files, it would be possible to revise individual program lines to adapt the recognition to the problem being analyzed. In this context, it would be worthwhile to investigate in future research how these discrepancies can be reduced, or even whether there are other systems that better perform the task of evaluating sexually suggestive poses. This issue could also be mitigated by enhancing the analytical capabilities of OpenPose. On the one hand, this can be done by incorporating the deeper version of this initialization into the pipeline rather than the demo version. The demo version seems to have problems recognizing certain postures. The full version method has not yet been implemented because the computer used to obtain the data does not have an NVIDA graphics card. For this reason, the demo version was first integrated into the pipeline, as no extensive technical requirements had to be met for this. On the other hand, the network must be specifically trained with such images, which were incorrectly recognized by the results or where the skeletons were only partially displayed. In this way, an attempt is made to improve the results in further research. The mentioned over-fitting of the neural network was characterized by an overtraining of the hand position. The results and the output image files have testified that many files were sorted into the posing category on basis of accentuated hand positions. The over-fitting should be reduced by using a much larger amount of data for further training and by setting the poses to be labeled more expressively. This adaptation should prevent the artificial intelligence from deciding whether a pose is sexually stimulating or not based on arm positions alone, but always acting in combination with the lower body region. The neural network was almost one hundred percent accurate in categorizing poses with spread legs into the *posing* category.

5.2 Work Restriction

The aforementioned classification problem can also be mentioned as a limitation of this work.Another limitation of this work is that a lack of available data sets hampered the analysis of the research objective. There was a lack of reliable and problem-specific material. Although the datasets used, each containing approximately 3,000 files, already contained less data than comparable datasets used to train other neural networks, the datasets used had to be further reduced. Deepening the dataset was not possible due to limited time, as the image files to be classified had to be labeled individually by hand. With the reduction of the data set, it had to be accepted that the neural network would not achieve the desired results. If the network would be trained with more data, a higher variability would arise and thus a higher probability would be given to recognize other poses. Furthermore, due to the legal situation, the main task, an improvement for the detection of child and youth pornography, could not be fully met. The developed pipeline has shown that sexually suggestive poses are classified into two categories *posing* and *no posing* with regard to the created workflow. However, since a one hundred percent transfer of the workflow to child pornographic files could not take place, only the recognition and marking of child skeletons was tested. This ran without problems. Based on this assumption, it can be assumed that the assumptions made here can also be transferred to the topic of posing in relation to children and adolescents, since everyday child skeletons were already recognized by the system.

6 Summary

The present study aimed to design awn worklow to improve the detection of child and adolescent pornographic files. To this end, insights from the field of motion analysis were to be used to distinguish posing images from everyday images based on classification. To answer the guiding question, a motion analysis system was connected to a neural network to draw conclusions about the valence of the data. It was found that OpenPose was the best motion analysis system of the systems tested to detect sexually suggestive poses. The meaningful results confirmed that the combination of the two models into a pipeline is suitable for detecting sexually suggestive poses of adults. In addition, the self-created neural network was integrated into the pipeline in such a way that it could be easily replaced based on the modification of a single program line. Thus, better and deeper neural networks could be inserted into the already working pipeline without a big effort. Thus, better and more accurate results could be achieved. Tests have also shown that the system can also be applied to the physique of children. For this reason, it can be assumed that the findings for sexually suggestive poses can also be transferred to the physique of children and adolescents. Thus, the basic goal of this thesis could be achieved and the research question could be confirmed. Furthermore, the recognition of posing can be used in other areas as well. For example, the learned knowledge from this work could be used to recognize intimacies between people, since elements of posing appear in many actions between individuals. For example, spreading the legs and positioning the buttocks in front of the camera occur more frequently in such scenarios. To subsequently implement this idea with the present pipeline, a more elaborate network would need to be created and trained with appropriate data. Afterwards, the model is saved and passed to the created pipeline in the form of a program line. In addition, it should be noted that the findings of this research alone cannot improve the analysis of child pornography files. A combination of several preliminary considerations must always be ensured to pursue this goal. For example, a network could be created that both detects posing and predicts certain sexually suggestive objects that describe a pornographic context. In this way, the data to be analyzed can be further minimized and the detection of incriminated files can be further improved.

Acknowledgments. At this point I would like to thank all those who have supported and motivated me during the preparation of this scientific work. First of all, I would like to thank my superior Nico Müller, who supervised and reviewed my bachelor thesis and this paper. I would like to thank him for giving me the opportunity to work on this tortious topic and to work on my own programming solutions and research. Thanks to him and my colleagues, my interest in digital forensics and AI-based solutions was awakened. Based on this new knowledge, I hope to gain further ground in this field. Finally, I would like to thank my family and friends who have always supported me in my scientific endeavors and were willing to face the theories that were created.

References

1. Mainzer, K. (2016). *Künstliche Intelligenz – Wann übernehmen die Maschinen?* (1. Aufl.). Springer. https://doi.org/10.1007/978-3-662-48453-1.
2. Cremers, A., Englader, et. al. (2019). Vertrauenswürdiger Einsatz von Künstlicher Intelligenz. Fraunhofer-Institut für Intelligente Analyseund Informationssysteme IAIS.
3. Schwander, M. Kultur neu Entdecken – SWR 2. Kinderpronographie im Netz- Neue Dimension des sexuellen Missbrauchs. (S. Striegl, & L. Meyer-Blankenburg, Redakteure) SWR (04. 08 2022). http://www.swr.de/swr2/wissen/kinderpornografie-im-netz-neuedimensionen-des-sexuellen-missbrauchs-swr2-wissen-2022-08-04-100.pdf 2 Okt. 2022.
4. Kuhnen, K. (2007). Was ist Kinderpronographie? 9, Hogrefe Verlag GmbH & Co. KG.Göttingen, https://doi.org/978-3-8017-2085-8.
5. Microsoft. Künstliche Intelligenz bewährt sich im Einsatz gegen Kinderpornografie., Microsoft (25. 05 2021). http://www.news.microsoft.com/de/kuenstliche-intelligenz-im-einsatz-gegen-kinderpornografie 27 Sept. 2022.
6. Garbers, N., & Brodthage, M. (2021). Herausforderungen beim Einsatz Künstlicher Intelligenz. In: Lecture Notes in Informatics (LNI), Informatik 2021, S. 879–889. G. f. e.V.
7. Kaplan, S., Handelmann, D., Handelmann, A. AI and Ethics. Sensitivity of neural networks to corruption of image classifcation, Springer Nature Switzerland AG., Schweiz S. 425–434. 6 Nov. 2020.

8. Banaeeyan, R., Karim, H. A., Lye, H., Fauzi, M. F. A., Mansor, S., & See, J. (2019). Automated nudity recognition using very deep residual learning network. *International Journal of Recent Technology and Engineering, 8*(3S), 136–141.
9. Moustafa, M. (2015). Applying deep learning to classify pornographic images and videos, arXiv:1511.08899.
10. Zecha, D., & Lienhart, R. Bestimmung intrazyklischer Phasengeschwindigkeiten, Universität Augsburg, Germany: Institut für Informatik, Augsburg, Juli 2022.
11. Xu, C., Xiaoye, L., Liya, M., & Xuan, F. AggPose: Deep Aggregation Vision Transformer for Infant Pose Estimation, Cornell University. 11 Mai 2022.
12. Brandl, C., Bonin, D., Mertens, A., Wischiniewski, S., & Schlick, C. M. Digitalisierungsansätze ergonomischer Analysen und Interventionen am Beispiel der markerlosen Erfassung von Körperhaltungen bei Arbeitstätigkeiten., Springer Verlag, Berlin Heidelberg., 27 Juli 2016, DOI: https://doi.org/10.1007/s41449-016-0016-9.
13. Eisele, J., & Franosch, R. (2016). Posing und der Begriff der Kinderpornografie in §184b StGB. *Zeitschrift für Internationale Strafrechtsdogmatik, 8*, 519–525.
14. Odemakinde, E. Estimation with Deep Learning – Ultimate Overview in 2022 http://www.viso.ai/deep-learning/pose-estimation-ultimate-overview/ 29 Sept. 2022.
15. Hao-Shu, F., Shuqin, X., et. all. RMPE: Regional Multi-Person Pose Estimation, arXiv:1612.00137v5, 04 Juli 2018.
16. He, K., Gkioyari, G., Dollar, P., Girshick, R. Mask R-CNN, arXiV:1703.06870v3, 24 Jan. 2018.
17. Sharma, S. A Short Guide to Pose Estimation in Computer Vision http://www.medium.com/@siddrrsh/a-short-guide-to-pose-estimation-in-computer-vision-3ea708dd9155, 02 Apr. 2020.
18. Cao, Z., Hidlago, G., Simon, T., Wei, S.-E., & Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. Cornell University. 10.48550. 30 Mai 2019.
19. Uhrmacher, L. Bachelorthesis. Gestenerkennung für den RoboCup mit OpenPose und neuonalen Netzen. Universität Hamburg. 29 Mai 2020.
20. Bösch, G. A Guide to OpenPose in (2022). viso.ai.http://www.viso.ai/deep-learning/openpose/. 29 Sept. 2022.
21. Figshare. http://www.figshare.com/articles/dataset/Adultcontentdataset/13456484/1. 20 Dez. 2020.

Human Centered Implementation Process of AI in SMEs – Conditions for Success

Carolin Böhme[1(✉)], Claudia Graf-Pfohl[2] and Katrin Meusinger[3]

[1] ICM – Institut Chemnitzer Maschinen- und Anlagenbau e. V., Otto-Schmerbach-Straße 19, 09117 Chemnitz, Germany
c.boehme@icm-chemnitz.de

[2] Technik und Bildung gGmbH, ATB Arbeit, Neefestraße 76, 09119 Chemnitz, Germany

[3] Silicon Saxony e. V., Manfred-von-Ardenne-Ring 20, Haus F, 01099 Dresden, Germany

Abstract. Artificial Intelligence (AI) is becoming an integral part of various aspects of human life. However, the successful implementation of AI systems poses significant challenges. Delays in the implementation of AI in Germany and Europe indicate hurdles, particularly for small and medium-sized enterprises (SMEs), which are important drivers of the German economy but also have structural disadvantages regarding AI implementation. To ensure that these AI systems are designed to meet the needs and expectations of end-users, Human Centered Design (HCD) has proven to be a promising approach. This paper aims to identify gaps and optimization potentials in the implementation process of AI with consideration of the Human Centered Design. The paper is based on existing literature and case studies to illustrate the benefits of using HCD in AI development and to identify weaknesses and optimization potentials in existing models. The paper concludes with recommendations for future research in this area.

Zusammenfassung. Die künstliche Intelligenz (KI) hat sich zunehmend in verschiedenen Bereichen des menschlichen Lebens etabliert. Jedoch geht die erfolgreiche Integration von KI-Systemen nicht ohne erhebliche Herausforderungen vonstatten. In Deutschland und Europa wurde die KI-Adoption durch Verzögerungen beeinträchtigt, besonders für kleine und mittelständische Unternehmen (KMU). Diese KMU bilden eine grundlegende Stütze der deutschen Wirtschaft, sehen sich jedoch spezifischen Hindernissen bei der KI-Einführung gegenüber. Ein vielversprechender Ansatz zur effektiven Ausgestaltung von KI-Systemen ist das Human Centered Design (HCD), welches den Menschen in den Vordergrund der KI-Entwicklung rückt. Dieser Artikel widmet sich der Analyse der KI-Einführung mit Schwerpunkt auf den Prinzipien des HCD, beleuchtet Defizite in vorhandenen Modellen und unterbreitet Empfehlungen für zukünftige Forschungsrichtungen. Das übergeordnete Ziel besteht darin, verantwortungsvolle und leistungsfähige KI-Systeme für diverse Anwendungsfelder zu fördern.

Die Implementierung von KI-basierten Assistenzsystemen oder KI-Anwendungen in Unternehmen steht häufig vor der Herausforderung, dass Mit-

C. Zinke-Wehlmann und J. Friedrich (Hrsg.): *First Working Conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow*, AIDRST 2023, Informatik aktuell, S. 83–99, 2024. https://doi.org/10.1007/978-3-658-43705-3_7

arbeiter mit Skepsis auf derartige Projekte reagieren. Um Akzeptanz und Partizipation zu steigern, erweist es sich als förderlich, Mitarbeiter bereits im frühen Stadium der Konzeption und Implementierung einzubinden. Des Weiteren ist es ratsam, einen Verhaltenskodex für die KI-Anwendung zu etablieren, um das Vertrauen der Mitarbeiter zu gewinnen. Eine hohe Beteiligung der Mitarbeiter geht einher mit gesteigertem Engagement und Akzeptanz. Die Entwicklung von Qualifikationen und Kompetenzmodellen muss den Anforderungen des digitalen Zeitalters gerecht werden. Eine zentrale Rolle spielen hierbei multimodales Lernen und Prozessunterstützung zur Förderung lebenslangen Lernens. Reflexionsschleifen tragen zur Steigerung der Lernerfolge und der Mitarbeiterleistung bei. Die Einbindung der Mitarbeiter fördert die Akzeptanz und aktive Teilnahme.

Zusammenfassend betont der Artikel die Bedeutung des HCD und die Notwendigkeit, Mitarbeiter frühzeitig in den Prozess einzubeziehen, um erfolgreiche und akzeptierte KI-Systeme zu gestalten. Er skizziert Empfehlungen für die praktische Umsetzung und für künftige Forschungsrichtungen, mit dem Ziel, KI-Implementierungen wirkungsvoller zu gestalten und das Potenzial dieser Technologie voll auszuschöpfen.

Keywords: AI · Human Centered Design · Implementation Process · SMEs

1 Introduction

Artificial Intelligence (AI) has become increasingly pervasive in various aspects of human life, ranging from healthcare and finance to education and entertainment. The integration and successful implementation of AI systems, however, present considerable challenges. In Germany and Europe, the adoption of AI has been hampered by delays and obstacles, with notable implications for small and medium-sized enterprises (SMEs) [1, 39, 16]. These SMEs play a crucial role as drivers of the German economy but face distinctive disadvantages in AI implementation. To ensure that AI systems are effectively designed to meet the needs of end-users, a promising approach known as Human Centered Design (HCD) has emerged. HCD focuses on placing human considerations at the forefront of AI development, thereby improving the overall user experience and system performance. By embracing a human-centered approach, potential gaps, and areas for optimization in the implementation process of AI can be identified. This paper aims to delve into the realm of AI implementation, with a particular focus on the application of HCD principles. Drawing upon existing literature and relevant case studies, the paper seeks to illustrate the benefits of adopting HCD in AI development while highlighting weaknesses and optimization potentials within existing models. By analyzing these gaps, it becomes possible to propose recommendations for future research in order to enhance the effectiveness and impact of AI implementation processes. The subsequent sections of this paper will provide an in-depth exploration of the challenges faced in implementing AI systems in Germany and Europe, with a specific emphasis on SMEs. Additionally, the principles and practices of HCD will be examined to shed light on its value in addressing these challenges. The paper will also present case studies to illustrate successful applications of HCD in AI development.

Finally, based on the insights gained from the analysis, recommendations will be put forth for future research directions in this area. By critically examining the implementation of AI and emphasizing the significance of HCD, this paper aims to contribute to the advancement of AI technologies that are efficient and aligned with the needs of end-users. Ultimately, this will foster the responsible and effective deployment of AI systems, leading to widespread benefits and advancements across various domains.

2 Methodology

To achieve the objectives of this paper, a systematic literature analysis was undertaken in March of 2023. A forward meta-analysis of papers was carried out, resulting in 500+ entries. The search was conducted using specific keywords related to HCD and AI implementation, namely "HCD+AI implementation." By employing this approach, a comprehensive collection of relevant scholarly articles, conference papers, and other scientific publications was gathered. The search results were then reviewed and filtered based on predefined research criteria. A specific backward search was carried including the terminology "human focus" and "SME" which narrowed down the available literature sources and opened up the research gap. The criteria furthermore ensured that the selected sources were directly related to the intersection of HCD and AI implementation with specific focus on the human and the SME sector, providing insights into the challenges and optimization potentials in this domain. In terms of ensuring quality and topicality (in-depth papers not older than 5 years) of the research, peer reviewed (quality assessment) and highly ranked (topicality) journals were used for the in-depth analysis. The intention of this systematic literature analysis was to gather a representative sample of the existing body of knowledge on the topic. An effort was made to cover a wide range of publications to capture the diversity of perspectives and findings in this field. Through the analysis of the selected studies, common weaknesses and gaps in existing models were identified. These overlaps in weaknesses provided evidence that the identified challenges were not isolated incidents but rather prevalent across different AI implementation models. This leads to the conclusion that the weaknesses identified in the examined studies are likely to exist in other models as well. Based on the findings from the literature analysis, recommendations were derived to address the identified gaps and optimize the AI implementation process. These recommendations serve as valuable insights for future research and development efforts in the field of HCD and AI implementation, aiming to enhance the design, deployment, and user experience of AI systems. It is important to note that this methodology of systematic literature analysis provided a rigorous and structured approach to gather and analyze existing scientific knowledge. The inclusion of multiple scientific databases and the use of specific search terms ensured a comprehensive search process, while the predefined research criteria guided the selection and examination of relevant sources.

3 State of the Art

HCD is a process-oriented approach, according to Norman [2013], which focuses on designing products, services, and systems tailored to the needs and requirements of users. The HCD process involves several interconnected steps that may include iterative loops:

1. Understanding users: Comprehending the needs, desires and behaviors of users.
2. Defining problems: Specification of the requirements and constraints that must be considered during development.
3. Ideation: Generating ideas to solve the defined problem.
4. Prototyping: Implementing ideas into tangible solutions.
5. Testing and Feedback: Testing prototypes with users to obtain reactions and feedback.
6. Refinement: Re-adjusting and adapting concepts based on results from the testing step [cf. 36].

The integration of Human-Centered Design into the developing process of AI is the result of an increasing focus on designing technology that aligns with user needs and preferences. The integration of HCD-approaches into the AI-implementation process enables faster and more efficient data processing, as well as generating insights that can improve the user experience and support decision-making [cf. 53].

3.1 Comparison of Existing Models

Overall, there are various ways in which HCD and AI can intersect. The choice of approach depends on specific requirements, including goals, target audience, and available resources. However, careful integration of AI into the design process can help create better products and services that align with the needs and perspectives of users. HCD is a crucial approach to ensure that the development and implementation of AI systems are aligned with the needs and values of humans. However, there is no one-size-fits-all approach for implementing HCD in AI, and various models have been proposed by researchers and practitioners. This topic aims to compare existing models for HCD in AI-implementation, explore their strengths and weaknesses, and identify potential areas for improvement. By analyzing and synthesizing these models, optimization potentials can be conducted, which can be considered in future research (Table 1).

This paper mainly focuses on the approach of human-centered AI design. The further theories shown are seen as complementary to the primary approach. As mentioned above, implementation strategies of AI projects often lack a human-centered approach and neglect the consideration of humans during the implementation process [cf. 43; 21]. However, there are SME criteria that should be considered when implementing AI projects, such as limited resources, limited expertise, and often a lack of experience in implementing AI projects [cf. 49]. Although there are good models HCD, they are often not implemented, leading to inadequate implementation of AI projects. Overall, the models are a useful contribution to the theoretical implementation of AI, but the implementation of real use cases is problematic.

Table 1. Comparison of different HCD-AI-Approaches

Approach	Description	Strengths	Weaknesses	Optimization potential
Human-Centered AI Design [cf. 4]	refers to integration of HCD principles into development of AI systems; AI systems must consider the needs, abilities, perspectives of users to ensure suitability for human use; helps to make AI systems more effective and user-friendly	increased user trust and acceptance; improved usability and accessibility; reduced bias and discrimination	requires significant expertise and resources; possibility of ethical trade-offs	integration of user feedback throughout the AI development process; ongoing evaluation and refinement of AI systems
AI-Supported Human-Centered Design [cf. 7]	assistance in carrying out tasks within scope of HCD; AI may help analyze user data and identify patterns to gain a better under-standing of user needs; AI may aid in generating design options and prototypes to accelerate and enhance design process	better efficiency and accuracy in data analysis; increased design innovation and creativity	dependent on data quality and availability; risk of over-reliance on AI-generated insights	development of transparent and interpretable AI models; collaboration to ensure balance between human and AI insights
AI-Driven Human-Centered Design [cf. 28]	AI as direct role in HCD by being incorporated as part of design itself; may be used for product/service personalization and recommending individual customization; may improve user experience by analyzing interactions between users and products/services and adjusting user needs and preferences	increased speed and scalability of design processes; ability to process large and complex data sets	possibility of reduced human input and creativity; risk of designing for the AI rather than for the user	integration of user feedback and HCD principles in the AI development process; regular evaluation of the impact of AI on the design process

(Fortsetzung)

Table 1. (Fortsetzung)

Approach	Description	Strengths	Weaknesses	Optimization potential
Ethics-Based Human-Centered Design [cf. 25]	integration of ethical considerations into HCD; ensure that AI systems are human-centered and ethically sound; includes consideration of multiple factors (privacy, discrimination, fairness in development of AI systems; may lead to creation of AI systems that are conform to ethical standards	higher trust and social responsibility; alignment with regulatory requirements and ethical standards	potential for conflicting ethical values and perspectives; need for expert guidance in ethical decision-making	integration of multiple perspectives and stakeholder input; ongoing evaluation of ethical implications throughout the design process

4 Soft and Missing Spots of Human-Centered AI Implementation

The introduction of AI in SMEs by using human-centered methods requires not only the involvement of humans in all considerations of technology introduction. Based on the indications of previous study, a socio-technical specification sheet proves to be effective as a starting point, in which the three dimensions of human-technology-organization are considered in the specific implementation project, here AI introduction [cf. 21; 40; 44; 50], as a human-centered design of the AI process cannot only focus on the consideration of human-technology interaction. The design of work processes by managers and organizational specifications must also be considered if AI is to be developed and used in a human-centered way [cf. 23].

4.1 Technology-Organization-Spots

The implementation of AI solutions in companies can bring new challenges. Weber et al. [cf. 51] assume that organizations are better positioned in this process if they have specific resources. These so-called organizational capabilities include AI project planning and development, data management, and AI model lifecycle management. The development of these abilities in the face of new and systemic challenges leads to unforeseeable processes in addition to the consideration of human-centered design in AI implementation. A study on user experience and usability among scientific developers showed that their orientation in the development of corresponding systems is not so much guided by user experience as by management and organizational issues [cf. 13]. For example, the successful implementation of AI in the education sector is recommended to be accompanied by the cooperation of researchers with local teachers so that organizational factors can be incorporated into the implementation process, which can affect the sustainable usability of AI [cf. 6]. Al Ali and Badi conclude that AI introductions should be accompanied by organization-wide change management processes because they can affect all employees and their work organization [cf. 2]. The question here is how technological and organizational processes can be integrated to design the introduction of AI as a supporting and – in some cases – disruptive technology. This presents SMEs in particular with additional challenges. SMEs should plan to implement the understanding of the AI implementation process as well as ensuring human-centeredness in day-to-day business [cf. 37]. Therefore, the approach of using an HCD analysis when implementing AI [cf. 53] is interesting. However, the effects of AI use are felt across departments and must be incorporated into business processes. Therefore, it is worth considering whether focusing on HCD regarding the directly affected users is sufficient or whether the involvement of proprietary stakeholders should be considered (e.g., managers, HR developers, [cf. 3]). Although the national AI strategy for Germany recommends responsible development and use of AI, the use of AI in Germany lags forecasts [cf. 18]. The SMEs predominant in Germany view the use of new technologies such as AI skeptically and rely on conventional technologies, such as rule-based systems [cf. 45]. A human-centered introduction of AI in German SMEs must also consider organizational restraints and the usage behavior of previous technologies to ensure a human-centered AI development process. However, it is also to be considered whether human-centered factors of the AI implementation process, as well as ethical considerations, could be implemented as dangers in the life cycle of an AI system [cf. 50].

4.2 Human-Technology-Spots

As stated, it is recommended to design the process of implementing AI with a human-centric approach. This also involves the human-centered design of human-machine interfaces as the connection between humans and technology in the company. For example, the ability to use AI is also a design issue, which can be ensured by early and transparent involvement of users in the development and design process of AI itself [cf. 31]. On the other hand, in the study of "Participatory Design in the Engineering Design Educational Environment," user involvement is not participatory in development, but tested through experiments afterward [cf. 12]. Technology acceptance also depends on whether the personal needs of users are considered, which includes process knowledge about the possibilities of AI and participation in defining the problem background [cf. 52; 38]. For example, it is part of the process to determine the scope of involvement of a worker together with the work planner and production manager for a human-centered introduction of AI [cf. 38]. However, this approach in practice faces limitations when workers cannot be released from the company's perspective (e.g., staff shortages) or when conflicts arise regarding the interpretation and legitimation of human-friendly work design [cf. 9]. The implementation of ethical and human-centered design guidelines must be evaluated and negotiated in corporate practice. Corresponding guidelines are defined as principles that must be translated into the daily work of employees, managers, and executives [cf. 37; 47]. Currently, there is still discussion about which competencies are even useful for human-centered AI use, which roles AI and humans take on, and how complex computational processes should be taught from a learning theory perspective [cf. 17; 35]. Regarding proprietary users (in the work system environment) in the company, such as executives and personnel managers, this knowledge is relevant to support workplace-related competency development processes. They can only become role models for digital transformation in the company if corresponding gaps are closed.

4.3 Human-Organization-Spots

However, the need for competency development among members of the organization extends not only to AI application knowledge regarding the technology solution to be introduced. Rather, the requirements for interdisciplinary skills associated with changes in work processes and responsibility structures are also increasing. A new leadership-, error-, and knowledge culture is considered essential in the AI implementation process to make the disruptive technology beneficially used in the company [cf. 21]. Regarding the roles negotiated in the socio-technical system between humans and the organization, the following requirements arise, which are inadequately considered in existing AI implementation models. Competency development is often implemented as AI training, but informal learning and creative processes in the team and workplace, and thus the integration of the company's learning process into the AI design process, are particularly important for adaptation and fit [cf. 32]. Employees and managers are equally affected by the transformation of the digital working world. For managers, it may become necessary to establish a new leadership culture [cf. 19]. Herrmann and Pfeiffer speak of organizational embedding of the AI implementation process when changed role requirements as well as rules for dealing with AI results and constraints must be considered in the everyday work of managers and HRM [cf. 2022]. In the area of AI implementation, there is increasing discussion about whether AI is attributed a social role. The use of ML algorithms as affective computer agents that interact with work groups through suggestions and participation in conversations, or that can proactively detect emotional and stress situations, is a growing topic within the team awareness discussion [cf. 48; 41]. The integration itself as well as the recommended methodology within the process of integrating employees into the competency development process in the context of AI applications is still unclear. The issues range from accounting aspects and the question of cost-effective integration of employees to the determination of performance indicators for tracking learning progress. Competence development as such remains a permanent topic in organizations, which is further reinforced by personnel shortages and lateral entrants. In addition, there is a backlog in the development of digital competences, especially in SMEs. In SMEs, qualification progress is tracked more manually and therefore not updated daily, and soft skills are only partially mapped [cf. 42; 5]. The measurement of knowledge and understanding of AI applications, especially in the competency development process of employees, is unclear [cf. 41]. Currently it cannot be adequately proven at what point an employee has sufficiently understood the AI to be able to apply, operate, and monitor it. The operationalization of experiential knowledge is not finally clarified [cf. 41]. Studies on digital skills in German SMEs also show that digital competencies are not being implemented mainly due to lack of time and high costs [cf. 34]. An evaluation by the OECD of 100 case studies on AI implementation, however, showed higher competency requirements that accompany the use of AI [cf. 33]. In addition, it becomes clear that the role of informal learning in the AI implementation process is often underrepresented, although it offers potential for human-centered adaptation of AI in the workplace and in the team. A shared definition of the role of AI in the organization and a clear operationalization of the competency development process are necessary to successfully implement AI [cf. 32]. In conclusion, the main challenges of implementing AI lie in digital skills, but also in the acceptance of organizational change. According to the HTO (human-technology-organization)-system [cf. 44] from a work design perspective, the role of humans remains unclear. This means that the relationship between humans and the organization is not co-developed – Human-Machine-Interface (human-technology) is considered as well as socio-technical systems (technology-organization), but not H-O (human-organization). This is especially problematic due to the overload of H-T (human-technology) design research before H-O design research, which risks prioritizing technological design over human-centered design, contradicting the principles of HCD.

5 Optimization Potentials and Solution Approaches

Building on the soft and missing spots in current models, this paper presents solutions that are currently being tested in German business practice within a practical AI project.

The Competence Center for the Future of Work in the "PerspektiveArbeit Lausitz" (PAL) project investigates how AI solutions can be implemented in a practice-oriented manner. The aim is to introduce data-based assistance systems, including artificial intelligence, in small and medium-sized enterprises in Lusatia in Saxony and Brandenburg in order to cope with structural change. This is a practical project because the AI solutions are ensured through participatory involvement of teams from four universities, 23 companies, and associations. The "PerspektiveArbeit Lausitz" project is being funded within the framework of the funding guideline "Future of Work: Regional Competence Centers for Labor Research. First round of competition: Design of new forms of work through artificial intelligence" in the program "Innovations for the production, service and work of tomorrow" by the Federal Ministry of Education and Research under the funding number 02L19C306 with a project duration of 01.11.2021 – 31.10.2026. The Lusatia region is characterized by three major challenges, the so-called "3D": demographic change, decarbonization and digitalization. Demographic change is reflected in Lusatia by a sharp decline in population and an aging population. The population of Lusatia fell by around 25% between 1990 and 2019, while the proportion of people over 65 rose from 15.5% to 26.1% [cf. 55]. This development also has an impact on the labor market, as there is a threat of labor shortages. Decarbonization is another challenge for Lusatia, as the region has traditionally been characterized by lignite mining and electricity generation. However, the energy transition requires a move away from coal and the expansion of renewable energies. This poses major challenges for the economy in the region and requires investments in new technologies and infrastructures [cf. 55]. Digitalization is another challenge for Lusatia, as the region has so far been less developed in this area than other regions in Germany. SMEs in the region, in particular, are struggling to meet the demands of digital transformation and need to catch up in this area [cf. 54]. These challenges require a targeted strategy and investments in the region to maintain competitiveness and create sustainable jobs. Based on the challenges presented by the "3D" in the Lusatia region, action needs to be taken to address these issues and ensure the region's competitiveness and sustainability. Deriving from the soft and missing spots in the AI implementation models and the experiences from the project PAL, four recommendations for action can be derived as to how successful AI implementation can be accompanied.

5.1 Overall: Reflection Loop

To ensure effective reflection and learning, it is important to not leave the HTO reflection loop until the end of a process, but instead reverse the process and incorporate reflection as an initial step [cf. 8]. As shown in the PAL-CRISP-DM model, an optimization approach of the CRISP-DM-model, it is possible to incorporate reflection loops throughout the whole process which meet the first optimization potential (see Fig. 1).

The introduction of data-based assistance systems or AI applications in companies often faces the challenge that employees are skeptical or even critical of such projects. In addition to the fear of losing their jobs due to technology, in many cases, technophobia plays a role, especially among older employees who are afraid of not being able to meet the resulting requirements. It is therefore imperative to involve the employees directly affected by the project in the conception and implementation of the project from the beginning and not only to "take them along" in the sense of "informing" them, but to involve them in the implementation of the project and to include their know-how. A first prelude for this can be an AI acceptance workshop in which, on the one hand, the legal framework conditions (for example, relevant aspects of the GDPR) and, on the other hand, the management's expectations of the project as well as a description of the project are discussed with the employees involved. Ideally, this workshop is designed to be low-threshold, for example, by first discussing the positive aspects of AI applications and examples of using AI solutions in everyday life.

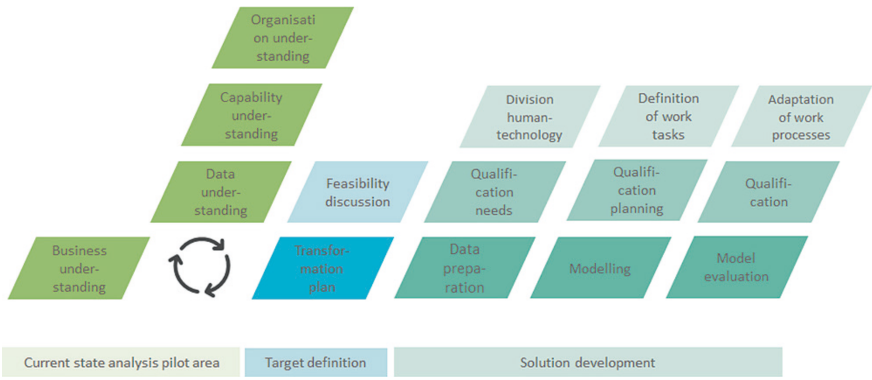

Fig. 1 AI procedure model [20, 10]

It is also helpful if the company agrees to a code of conduct for the application of AI in this context. For example, the "Human Friendly Automation Value Manifesto" is a suitable option. Such a code of conduct can visibly communicate the company's approach to the application of AI both internally and externally, thereby providing employees and potential applicants with a trustworthy and secure interaction with the new technologies. Such an AI workshop should then also be conducted with other employees in the next step. This way, uncertainties and resulting unrest in the workforce, which becomes aware of the project but has no detailed information about it, can be avoided early on. The main message of such a workshop, to which the company management must also position itself clearly, should be: "We are shaping this process together!"

5.2 Human-Organization: Competence Development Needs Misjudged

To meet the challenges posed by the changing demands of the digital age, it is necessary to develop new qualifications and competence development models. Moving away from individual training to the integration of multimodal learning approaches, as well as process support, is crucial for employees to remain willing to engage in lifelong learning. According to a study by the European Commission, learning in the workplace is crucial for employees to adapt to changing requirements and maintain their employability [cf. 15]. To achieve success, non-formal and informal learning must be planned and structured directly with the help of team processes and reflection loops, with iterative reflection loops being a key factor. Research shows that reflection loops can be an effective tool for improving learning outcomes and enhancing employee performance [cf. 29]. This involves considering not only technological feasibility but also the design of informal learning opportunities. Process thinking is essential for employees to move towards lifelong learning. However, digitalization in public administration can be hindered by factors such as process chains from the 1970s, which cannot be solved through training. Therefore, process support and project management are required to help employees overcome such obstacles in both the market and the workplace [cf. 24]. Job rotation is a potential solution to help employees adapt to changing demands. Studies have shown that job rotation can have a positive impact on employees' skill development, job satisfaction, and career advancement [cf. 46]. Furthermore, the example of introducing AI-assisted systems highlights the importance of planning the timing of competency development. Although these systems can demonstrate their functionality and benefits through use and application, the understanding of the processes must be established beforehand. Otherwise, an informed decision to participate in the adaptation of the AI solution cannot be presented. This means that necessary competence development steps must already be carried out in the company before the test phase, so that the employees are empowered sufficiently. Research suggests that involving employees in the development and implementation of AI solutions can increase their engagement and commitment to the technology [cf. 26]. Within the framework of the project PAL, a participatory approach is adopted in project and process management, with an initial workshop involving employees, managers, and scientists to address any initial questions, create acceptance, and establish initial process understanding, especially for SME [cf. 21]. Including the (learning) experience of users in the implementation process addresses all optimization potentials of AI approaches.

5.3 Human-Organization: Termination Criteria and Inclusion

In the implementation of AI systems, it is of great importance to define termination criteria in order to minimize possible errors and risks. Recommendations for discontinuation criteria can be based on, for example, error rates, trustworthiness of the data and the application, ethical considerations, and regulatory requirements. It is important to strike a balance between avoiding errors and the benefits of the application. Too restrictive an approach can limit the benefits of AI systems, while too open an approach increases the risk of errors and harm. Terminations criteria must be defined at the beginning of each process. For instance, agile lab teams can be defined, that accompany the process on a permanent basis and are responsible for checking the termination and success criteria regularly [cf. 19]. This supports transparency and collaboration between developers and users, as called for in the second and third optimization potential. The inclusion of employees in the process should also be considered in order to achieve a high level of acceptance and participation. This can be done through training, workshops, and participatory approaches. A high level of employee participation can help identify and address concerns and challenges early on. Active involvement can also promote employee acceptance and motivation, which in turn supports the success of the implementation. It is important to note, however, that employee inclusion should not be seen only as a means to an end. The inclusion of employees in the process should be pursued as a goal in its own right, as this can contribute to a more equitable and inclusive working environment [cf. 11].

5.4 Technology-Organization: Further Development of Success Indicators

In addition to Key Performance Indicators (KPIs) that measure the number of loops and areas of congestion, it is also important to use satisfaction scores as a measure of the success of AI implementation in a company. Customer and employee satisfaction are crucial in ensuring the effectiveness of the AI system and can provide valuable insights for further improvement [cf. 22]. AI can help improve customer experience and satisfaction by providing personalized recommendations, faster response times, and more efficient service. This can be measured through customer feedback surveys and ratings. Furthermore, AI can help improve employee satisfaction by reducing mundane tasks, increasing autonomy, and providing better decision-making support. This can be measured through employee feedback surveys and retention rates [cf. 14]. Companies should aim to empower themselves to make further adjustments to the system without relying on external support. This involves a deep understanding of the implemented AI technology and developing internal expertise for system maintenance and improvement. As part of the company-specific transformation plan (see Fig. 1), satisfaction metrics support the evaluation of the project, which addresses the fourth optimization potential

6 Conclusion and Outlook

Considering the theoretical considerations, the view on the implementation figures of AI in German SMEs, and the practical application in the PAL project, a human-centered design for AI implementation can be complemented as following: Human-centered design considers all persons affected by the product, including executives and personnel developers beyond the boundaries of the work system [cf. 9]. Furthermore, the connection between HCD and HTO is beneficial for practice, as it reveals gaps and solutions in the development of Human-Technology design. However, the practical relevance of HCD has not been adequately evaluated. For instance, what constitutes sufficient involvement of affected parties? Is feedback from users sufficient, or should proprietarily affected parties, such as executives and leaders of other departments, be included? Therefore, a participatory approach should be adopted from the outset to incorporate practical use cases. Particularly, the neglect of reservations and technology experience of executives and management should be considered in AI implementation processes. A human-centered AI implementation is always an intervention in the prevailing work system and therefore organizational processes. Especially in the Competence Centers for Labor Research in Germany, which focus on the design of the use of methods and tools of artificial intelligence, there is a chance of human-centered design of future work. At the same time, the requirement for a human-centered development and introduction of AI requires caution with the obstacles highlighted. Especially for SMEs, this offers the opportunity to try out the use of AI and reduce the digital divide through AI [cf. 27]. This opportunity is particularly important in view of the unclear risk assessments of AI use, in order to open up usage perspectives for SMEs [cf. 30]. The strength of the contribution is the consistent HCD along established socio-technical dimensions. The solution approaches of AI-Implementation should be further developed [cf. 8].

References

1. Accenture. (2022). The art of AI maturity: Advancing from practice to performance. https://www.accenture.com/_acnmedia/Thought-Leadership-Assets/PDF-5/Accenture-Art-ofAI-Maturity-Report.pdf.
2. Al Ali, A., & Badi, S. (2022). Exploring the impacts of Artificial Intelligence (AI) implementation at individual and team levels: A case study in the UAE government sector. In M. Themistocleous & M. Papadaki (Hrsg.), *Information Systems. EMCIS 2021. Lecture notes in business information processing* (437). Springer, Cham. https://doi.org/10.1007/978-3-030-95947-0_42.
3. Arslan, A., Cooper, C., Khan, Z., Golgeci, I., & Ali, I. (2022). Artificial intelligence and human workers interaction at team level: A conceptual assessment of the challenges and potential HRM strategies. *International Journal of Manpower, 43*(1), 75–88.
4. Auernhammer, J. (2020). Human-centered AI: The role of Human-centered Design Research in the development of AI. In S. Boess, M. Cheung, & R. Cain (Hrsg.), *Synergy – DRS International Conference 2020*, 11–14 August, Held online. https://doi.org/10.21606/drs.2020.282.

5. Bartelheimer, P., Domnick, F., & Siemens, J. (2019). *Digitalisierung im Mittelstand: Bestandsaufnahme und Handlungsempfehlungen*. Institut der deutschen Wirtschaft Köln (IW).
6. Bhimdiwala, A., Neri, R. C., & Gomez, L. M. (2022). Advancing the Design and Implementation of Artificial Intelligence in Education through Continuous Improvement. *International Journal of Artificial Intelligence in Education, 32*, 756–782. https://doi.org/10.1007/s40593-021-00278-8
7. Beck, M., & Libert, B. (2017). *The rise of AI makes emotional intelligence more important*. Harvard Business Review online https://hbr.org/2017/02/the-rise-of-ai-makes-emotional-intelligence-more-important.
8. Bendel, A., & Latniak, E. (2023). Weiter so mit MTO? Konzeptionelle Entwicklungsbedarfe soziotechnischer Arbeits- und Systemgestaltung. *Gr Interakt Org, 54*, 9–26. https://doi.org/10.1007/s11612-023-00669-6
9. Braun, M. (2022). Menschengerechte Gestaltung KI-gestützter Arbeitssysteme. *Betriebliche Prävention, 5*, 4. https://doi.org/10.37307/j.2365-7634.2022.05.04.
10. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*. The CRISP-DM consortium.
11. Cheng, C., & Liang, J. (2019). A participatory design approach to developing a knowledge-based system for the home-based health care industry. *Applied ergonomics, 77*, 146–156.
12. Chien, Y. H., Hsiao, H. S., Chang, Y. S., & Yao, C. K. (2020). The development of an artificial intelligence assistant for participatory design in the engineering design educational environment. In M. Auer, H. Hortsch, & P. Sethakul (Hrsg.), *The Impact of the 4th Industrial Revolution on Engineering Education. ICL 2019. Advances in Intelligent Systems and Computing*, (1134). Springer, Cham. https://doi.org/10.1007/978-3-030-40274-7_36.
13. Clemmensen, T., Iivari, N., Rajanen, D., & Sivaji, A. (2022). "Organized UX Professionalism" – An empirical study and conceptual tool for scrutinizing UX work of the future. In , et al. *Human work interaction design. Artificial intelligence and designing for a positive work experience in a low desire society. HWID 2021. IFIP Advances in Information and Communication Technology* (609). Springer, Cham. https://doi.org/10.1007/978-3-031-02904-2_2.
14. Cubric, M. (2020). Drivers, barriers and social considerations for AI adoption in business and management: A tertiary study, Technology in Society, Volume 62. *ISSN, 101257*, 0160-791X. https://doi.org/10.1016/j.techsoc.2020.101257
15. European Commission. (2016). Study on the impact of learning mobility on the development of professional competences. https://op.europa.eu/en/publication-detail/-/publication/56d7a8ed-2af5-11e6-b5ec-01aa75ed71a1.
16. Federal Ministry for Economic Affairs and Energy. (2019). SMEs digital. Strategies for the digital transformation. https://www.bmwk.de/Redaktion/EN/Publikationen/Mittelstand/smesdigital-strategies-for-digital-transformation.pdf?__blob=publicationFile&v=5.
17. Gašević, D., Siemens, G., & Sadiq, S. (2023). Empowering learners for the age of artificial intelligence. *Computers and Education: Artificial Intelligence, 100130*. https://doi.org/10.1016/j.caeai.2023.100130.
18. Giering, O. (2022). Künstliche Intelligenz und Arbeit: Betrachtungen zwischen Prognose und betrieblicher Realität. *Zeitschrift für Arbeitswissenschaft, 76*(1), 50–64. https://doi.org/10.1007/s41449-021-00289-0
19. Graf-Pfohl, C. (2022). Agilität als zukunftssicheres Merkmal digitaler Führung. In G. Lanza, P. Nieken, P. Nyhuis, & A. Trübswetter (Hrsg.), *Digitale Führung und Technologien für die Teaminteraktion von morgen. Praxisbezogene Methoden und Anwendungsbeispiele* (S. 66–81). TEWISS-Technik und Wissen, Garbsen.
20. Gröllich, D., Hahmann, M., Ott, G., & Graf-Pfohl, C. (2023). Vorgehensweise zur Implementation von datenbasierten Assistenzsystemen. In H. Mittweida (Hrsg.), *Workshop Entwicklung hybrider Arbeitssysteme* (S. 35–38). Hochschule Mittweida, Mittweida.

21. Herrmann, T., & Pfeiffer, S. (2022). Keeping the organization in the loop: a socio-technical extension of human centered artificial intelligence. *AI & Society, 38,* 1523–1542. https://doi.org/10.1007/s00146-022-01391-5
22. Hildebrandt, T., Fettke, P., & Loos, P. (2019). Unterstützung des Prozessmanagements durch Künstliche Intelligenz. In Jan vom Brocke (Hrsg.), *Handbuch Prozessmanagement* (2. Aufl., S. 1–24). Springer Berlin Heidelberg.
23. Huchler, N. (2022). Komplementäre Arbeitsgestaltung. Grundrisse eines Konzepts zur Humanisierung der Arbeit mit KI. *Zeitschrift für Arbeitswissenschaft, 76*(2), 158–175. https://doi.org/10.1007/s41449-022-00319-5.
24. Huchler, N. (2022). Digitalisierung von Verwaltungsprozessen. In A. Homburg, S. Bohnet-Joschko, & S. J. Schmidt (Hrsg.), *Handbuch Digitalisierung in Staat und Verwaltung* (S. 99–115). Springer.
25. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2017). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems, version 2*. IEEE. http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html.
26. Kerzel, M., Koch, M., & Schumacher, C. (2021). Participatory design of AI-assisted decision-making systems in public administration. In Proceedings of the 54th Hawaii International Conference on System Sciences, (S. 4281–4290).
27. Kitsara, I. (2022). Artificial intelligence and the digital divide: From an innovation perspective. In A. Bounfour (Hrsg.), *Platforms and artificial intelligence. Progress in IS* (pp. 245–265). Springer, Cham. https://doi.org/10.1007/978-3-030-90192-9_12.
28. Kussmaul, R., Biedermann, M., Pappas, G., Jónasson, J., Winiger, P., Zogg, M., & Ermanni, P. (2019). Individualized lightweight structures for biomedical applications using additive manufacturing and carbon fiber patched composites. *Design Science, 5*, E20. https://doi.org/10.1017/dsj.2019.19
29. Lian, R. J., & Chen, C. H. (2014). Enhancing team reflection and performance: The roles of feedback and reflection modes. *Journal of Business and Psychology, 29*(1), 113–129. https://doi.org/10.1007/s10869-013-9302-2
30. Liebl, A., & Klein, T. (2023). *AI act: Risk classification of AI systems from a practical perspective. A study to identify uncertainties of AI users based on the risk classification of more than 100 AI systems in enterprise functions*. Whitepaper, Munich: applied AI Initiative.
31. Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, (S. 1–16). https://doi.org/10.1145/3313831.3376727.
32. Long, D., Blunt, T., & Magerko, B. (2021). Co-Designing AI Literacy Exhibits for Informal Learning Spaces. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW2), 1–35. https://doi.org/10.1145/3476034
33. Milanzes, A. (2023). *The impact of AI on the workplace: Evidence from OECD case studies of AI implementation* (OECD Social, Employment and Migration Working Papers Nr. 289; OECD Social, Employment and Migration Working Papers, Bd. 289). https://doi.org/10.1787/2247ce58-en.
34. Müller, M. (2023). Skills shortage marks a turning point: The times of guaranteed growth are over. *KfW Research Focus on Economics, 414,* 1–10.
35. Ng, D. T. K., Leung, J. K. L., Su, M. J., Yim, I. H. Y., Qiao, M. S., & Chu, S. K. W. (2022). AI literacy on human-centered considerations. In *AI Literacy in K-16 Classrooms* (pp. 119–129). Springer, Cham. https://doi.org/10.1007/978-3-031-18880-0_9.
36. Norman, D. A. (2013). *The design of everyday things: Revised and expanded edition*. Basic Books. MIT Press; 2nd rev. and exp. ed. Edition. New York.

37. OECD. (2021). *The digital transformation of SMEs*. OECD. https://doi.org/10.1787/bdb9256a-en.
38. Pokorni, B., Zwerina, J., & Hämmerle, M. (2020). Human-centered design approach for manufacturing assistance systems based on Design Sprints. *Procedia CIRP, 91*, 312–318. https://doi.org/10.1016/j.procir.2020.02.181
39. Rammer, C., Fernández, G. P., & Czarnitzki, D. (2022). Artificial intelligence and industrial innovation: Evidence from German firm-level data. *Research Policy, 51*(7), 104555. https://doi.org/10.1016/j.respol.2022.104555
40. Roth, G., & Kötter, W. (2020). Soziotechnisches Lastenheft: Herangehensweise bei der Anforderungsermittlung an digitale Assistenzsysteme am Beispiel eines mittelständisch geprägten Industrieunternehmens. In D. GfA (Hrsg.), *Frühjahrskongress 2020, Berlin Beitrag B.5.4 Digitaler Wandel, digitale Arbeit, digitaler Mensch?* (pp. 1–5). GfA-Press, Dortmund.
41. Schaefer, K. E., et al. (2021). Human-autonomy teaming for the tactical edge: The importance of humans in artificial intelligence research and development. In W. F. Lawless, R. Mittu, D. A. Sofge, T. Shortell, & T. A. McDermott (Hrsg.), *Systems engineering and artificial intelligence* (pp. 115–148). Springer, Cham. https://doi.org/10.1007/978-3-030-77283-3_7.
42. Schröder, A., & Wenzel, S. (2020). Digitalisierung und Personalentwicklung in KMU. In *KMU: Status quo, Herausforderungen und Handlungsfelder. Eine empirische Analyse in Hessen*. Kompetenzzentrum für Fachkräftesicherung (KOFA).
43. Shneiderman, B. (2022). *Human-centred AI*. Oxford University Press. https://doi.org/10.1093/oso/9780192845290.001.0001.
44. Ulich, E. (2013). Arbeitssysteme als soziotechnische Systeme – eine Erinnerung. *Psychology of Everyday Activity, 6*(1), 4–12.
45. Ulrich, P., & Frank, V. (2021). Relevance and Adoption of AI technologies in German SMEs – Results from Survey-Based Research. *Procedia Computer Science, 192*, 2152–2159. https://doi.org/10.1016/j.procs.2021.08.228
46. Van der Heijden, B. I. J. M., Scholarios, D., Bozionelos, N., & van der Schoot, E. (2019). Job rotation: A review and research agenda. *Journal of Vocational Behavior, 110*, 416–433. https://doi.org/10.1016/j.jvb.2018.11.006
47. Vieweg, S. H. (2021). Ethical AI implementation. In S. H. Vieweg (Hrsg), *AI for the good. Management for professionals*. Springer, Cham. https://doi.org/10.1007/978-3-030-66913-3_11
48. Vijh, G., Sharma, R., & Agrawal, S. (2021). The heartfelt and thoughtful rulers of the world: AI implementation in HR. In P. K. Singh, G. Veselov, V. Vyatkin, A. Pljonkin, J. M. Dodero, & Y. Kumar (Hrsg.), *Futuristic Trends in Network and Communication Technologies. FTNCT 2020. Communications in Computer and Information Science* (1395). Springer, Singapore. https://doi.org/10.1007/978-981-16-1480-4_24.
49. Von Garrel, J., & Jahn, C. (2022). Design Framework for the Implementation of AI-based (Service) Business Models for Small and Medium-sized Manufacturing Enterprises. *Journal of the Knowledge Economy*. https://doi.org/10.1007/s13132-022-01003-z
50. Warnecke, H.-J. (1992). *Mensch, Technik, Organisation – Durch Umdenken zu neuer Wettbewerbsfähigkeit. M und E Forum 1992*. https://publica.fraunhofer.de/handle/publica/320002.
51. Weber, M., Engert, M., Schaffer, N., et al. (2022). Organizational Capabilities for AI Implementation—Coping with Inscrutability and Data Dependency in AI. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-022-10297-y
52. Weitz, K., Schlagowski, R., & André, E. (2021). Demystifying artificial intelligence for end-users: Findings from a participatory machine learning show. In S. Edelkamp, R. Möller,

& E. Rueckert (Hrsg.), *Advances in Artificial Intelligence. KI 2021. Lecture Notes in Computer Science* (12873). Springer, Cham. https://doi.org/10.1007/978-3-030-87626-5_19.
53. Zhu, W., Li, L., & Yang, L. (2018). A review of human-centered design for artificial intelligence. *Journal of Intelligent Manufacturing, 29*(2), 251–268.
54. Seibert, H., & Wapler, R. (2020). *Einwanderung nach Deutschland: Viele Hochqualifizierte, aber auch viele Ungelernte* (IAB-Kurzbericht 08/2020), Nürnberg (S. 12).
55. Brixy, U., Egeln, J., Gottschalk, S., & Kohaut, S. (2021). *Sonderbefragung des IAB-ZEW-Gründungspanels: Junge Unternehmen spüren die Corona-Krise stärker* (IAB-Kurzbericht 04/2021), Nürnberg (S. 8).

AI supported knowledge management and learning

Abstract. How can knowledge management and learning processes can be supported or enhanced by AI. In this session, the authors present approaches to the use of various AI tools and techniques that can help to organize, share and use knowledge efficiently. The session also addresses challenges and potentials of AI-supported knowledge management and how this can contribute to lifelong learning.

LLM-assisted Knowledge Graph Engineering: Experiments with ChatGPT

Lars-Peter Meyer[1,2,3(✉)], Claus Stadler[1,2,3], Johannes Frey[1,2,3], Norman Radtke[1,2], Kurt Junghanns[1,2], Roy Meissner[2,3], Gordian Dziwis[1], Kirill Bulert[1,2] and Michael Martin[1,2]

[1] Institut für Angewandte Informatik (InfAI) e. V., Leipzig, Germany
{lpmeyer,stadler,radtke,junghanns,dziwis,bulert,martin}@infai.org;
frey@informatik.uni-leipzig.de
https://www.infai.org,
[2] AKSW research group, Leipzig, Germany
roy.meissner@uni-leipzig.de
https://aksw.org,
[3] Erziehungswiss. Fakultät, Allgemeine Pädagogik, Universität Leipzig, Leipzig, Germany
https://www.uni-leipzig.de

Abstract. Knowledge Graphs (KG) provide us with a structured, flexible, transparent, cross-system, and collaborative way of organizing our knowledge and data across various domains in society and industrial as well as scientific disciplines. KGs surpass any other form of representation in terms of effectiveness. However, Knowledge Graph Engineering (KGE) requires in-depth experiences of graph structures, web technologies, existing models and vocabularies, rule sets, logic, as well as best practices. It also demands a significant amount of work.

Considering the advancements in large language models (LLMs) and their interfaces and applications in recent years, we have conducted comprehensive experiments with ChatGPT to explore its potential in supporting KGE. In this paper, we present a selection of these experiments and their results to demonstrate how ChatGPT can assist us in the development and management of KGs.

Zusammenfassung. Wissensgraphen (englisch *Knowledge Graphs*, KGs), bieten uns eine strukturierte, flexible, transparente, systemübergreifende und kollaborative Möglichkeit, unser Wissen und unsere Daten über verschiedene Bereiche der Gesellschaft und der industriellen sowie wissenschaftlichen Disziplinen hinweg zu organisieren. KGs übertreffen jede andere Form der Repräsentation in Bezug auf die Effektivität. Die Entwicklung von Wissensgraphen (englisch *Knowledge Graph Engineering*, KGE) erfordert jedoch fundierte Erfahrungen mit Graphstrukturen, Webtechnologien, bestehenden Modellen und Vokabularen, Regelwerken, Logik sowie Best Practices. Es erfordert auch einen erheblichen Arbeitsaufwand.

C. Zinke-Wehlmann und J. Friedrich (Hrsg.), *First Working Conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow*, AIDRST 2023, Informatik aktuell, S. 103–115, https://doi.org/10.1007/978-3-658-43705-3_8

In Anbetracht der Fortschritte bei groen Sprachmodellen (englisch *Large Language Modells*, LLMs) und ihren Schnittstellen und Anwendungen in den letzten Jahren haben wir umfassende Experimente mit ChatGPT durchgeführt, um sein Potenzial zur Unterstützung von KGE zu untersuchen. In diesem Artikel stellen wir eine Auswahl dieser Experimente und ihre Ergebnisse vor, um zu zeigen, wie ChatGPT uns bei der Entwicklung und Verwaltung von KGs unterstützen kann.

Key words: ChatGPT · knowledge graph engineering · RDF · large language model use cases · AI application

1 Introduction

In the last years, Artificial Intelligence (AI) has shown great promise in improving or revolutionizing various fields of research and practice, including knowledge engineering. The recent big leap in AI-based assistant chatbots, like ChatGPT (Generative Pre-trained Transformer) model, has created new opportunities to automate knowledge engineering tasks and reduce the workload on human experts. With the growing volume of information in different fields, the need for scalable and efficient methods to manage and extract knowledge from data that also adapt to new sources is critical. Despite the advances in research w.r.t. (semi)automation, knowledge engineering tasks still rely vastly on human experts. On one hand, this process can be time-consuming, resource-intensive, and susceptible to errors. On the other hand, the reliance on human expertise in knowledge engineering exposes it to workforce shortages (as knowledge engineers are scarce and the demand is growing) and the risk of expertise loss. These factors can impact the resilience and sustainability of systems and operations that rely on knowledge engineering. AI-based assistant bot approaches, such as ChatGPT, could bridge this gap by providing a unified tool for tasks in knowledge engineering, to reduce the workload of knowledge engineers themselves, but also make knowledge engineering more accessible to a broader audience. ChatGPT, in particular, has shown promise in generating responses in a variety of syntactical representations (including code and markup languages) to user queries or task descriptions written in natural language.

In this paper, we discuss and investigate the potential of ChatGPT to support or automate various knowledge engineering tasks (e.g. ontology generation, SPARQL query generation). We will explore the benefits, pitfalls and challenges of using it and identify potential avenues for future research.

2 Related Work

ChatGPT, a Large Language Model (LLM) published by OpenAI[1], raised the interest in the broad field of Machine Learning (ML)[2] and especially LLMs [4] on

[1] https://openai.com/blog/chatgpt

[2] https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf

a broad scale. While there are current discussions and analysis on the capabilities of LLMs like ChatGPT in general (e.g. [1]), there is little in the area of knowledge graph engineering. Ekaputra et al. [3] gives a general overview of current research on the combination of the broad field of ML and semantic web.

Searching Google Scholar and Semantic Scholar with “knowledge graph ChatGPT”, “ontology ChatGPT” and “rdf ChatGPT” in the beginning of April 2023 results in only two relevant papers. The first one, [7], reviews the differences between conversational AI models, prominent ChatGPT, and state-of-the-art question-answering systems for knowledge graphs. In their survey and experiments, they detect capabilities of their used frameworks but highlight ChatGPTs explainability and robustness. The second one, [6], discusses the usage of ChatGPT for database management tasks when tabular schema is expressed in a natural language. They conclude among others that ChatGPT is able to assist in complex semantic integration and table joins to simplify database management and enhance productivity. The applied approaches and results of these two papers indicate that the idea of using LLMs like ChatGPT in the field of KG engineering is encouraging and that the LLMs might assist KG engineers in their workflows. Still, the research on the usage of LLMs for knowledge graph engineers is scarce and seems to be a new research area.

There exist some non- and semi-scientific resources which render the topic from a practical and experience perspective. We want to highlight here a helpful blog post by Kurt Cagle [2] on ChatGPT for “knowledge graph workers” and a blog post by Konrad Kaliciński [5] on knowledgegraph generation in Neo4J assisted by ChatGPT.

3 LLM-Assisted Knowledge Graph Engineering – Potential Application Areas

In discussion rounds with knowledge graph engineering experts we identified the following preliminary list of potential use cases in the domain of knowledge graph engineering applicable to LLMs assistance:

- Assistance in knowledge graph usage:
  - Generate SPARQL queries from natural language questions (related experiment in Sect. 4.1 and 4.3)
  - Exploration and summarization of existing knowledge graphs (related experiment in Sect. 4.5)
  - Conversion of competency questions to SPARQL queries
  - Code generation or configuration of tool(chain)s for data pipelines
- Assistance in knowledge graph construction
  - Populating knowledge graphs (related experiment in Sect. 4.4) and vice versa
  - Creation or enrichment of knowledge graph schemas / ontologies
  - Get hints for problematic graph design by analysing ChatGPT usages problems with a knowledge graph

- Semantic search for concepts or properties defined in other already existing knowledge graphs
- Creation and adjustment of knowledge graphs based on competency questions

Given the limited space of this paper, we evaluate a subset of the application areas with experiments in the following section.

4 Experiments

To evaluate the capabilities of LLMs at the example of ChatGPT for assisting with knowledge graph engineering, we present several experiments and their results. Further details about them is given in the Supplemental Online Resources. Most experiments were conducted with ChatGPT with the LLM GPT-3.5-turbo[3] (named *ChatGPT-3* from here on), some additionally with ChatGPT with the LLM GPT-4[4] (named *ChatGPT-4* from here on).

```
:anne a foaf:Person ; foaf:firstName "Anne" ; foaf:surname "Miller" ;
  vcard:hasAddress [ a vcard:Home ; vcard:country-name "UK" ] .
:bob a foaf:Person ; foaf:firstName "Bob" ; foaf:surname "Tanner" ;
  vcard:hasAddress [ a vcard:Home ; vcard:country-name "US" ] .
:wonderOrg a org:Organization .
:researchDep a org:OrganizationalUnit ; org:unitOf :wonderOrg ;
  rdfs:label "Research Department" .
:marketingDep a org:OrganizationalUnit ; org:unitOf :wonderOrg ;
  rdfs:label "Marketing Department" .
:chiefResearchOfficer a org:Role . :marketingManager a org:Role .
[ a org:Membership ; org:member :anne ; org:organization :researchDep ;
  org:role :chiefResearchOfficer ] .
[ a org:Membership ; org:member :bob  ; org:organization :marketingDep ;
  org:role :marketingManager ] .
```

Listing 1: An organizational KG with two people working in different departments of the same organization.

4.1 SPARQL Query Generation for a Custom Small Knowledge Graph

For a first evaluation, we designed a small knowledge graph as shown in Listing 1. Specifically, we wanted to know whether (1) GPT can explain connections between indirectly related entities, (2) create SPARQL queries over the given model and (3) reconstruct the model if all properties and classes were relabelled.

We issued the following prompt, which includes the knowledge graph from Listing 1, on ChatGPT-3 and ChatGPT-4:

[3] https://platform.openai.com/docs/models/gpt-3-5
[4] https://platform.openai.com/docs/models/gpt-4

> **Prompt 1:** Given the RDF/Turtle model below, are there any connections between US and UK? <rdf-model>

In the knowledge graph of Listing 1, there is a connection between the two countries via the two people living in these, which got a job in different departments of the same company. While ChatGPT-3 fails to identify this relation, ChatGPT-4 successfully identifies it in all cases.

We further asked both ChatGPT models with prompt 2 and received five SPARQL queries each, which we analysed for their syntactic correctness, plausible query structure, and result quality. The results for prompt 2 are listed in Tab. 1 and show that both models produce syntactically correct queries, which in most cases are plausible and produce corrects results in 3/5 (ChatGPT3) and 2/5 (ChatGPT4) cases.

> **Prompt 2:** Given the RDF/Turtle model below, create a SPARQL query that lists for every person the country, company and department and role. Please adhere strictly to the given model. <rdf-model>

Table 1. Findings in generated SPARQL queries for prompt 2

	ChatGPT-3	ChatGPT-4
syntactically correct	5/5	5/5
plausible query structure	4/5	3/5
producing correct result	3/5	2/5
using only defined classes and properties	3/5	4/5
correct usage of classes and properties	5/5	5/5
correct prefix for the graph	5/5	4/5

In essence, AI-based query generation is possible and it can produce valid queries. However, the process needs result validation in two dimensions: 1) validating the query itself by matching to static information, like available classes and properties in the graph, as well as 2) validating the executed query results to let ChatGPT generate new queries in case of empty result sets in order to find working queries in a try & error approach.

As a last prompt on the knowledge graph from Listing 1, we created a derived RDF graph by relabelling all classes and properties with sequentially numbered IRIs in the example namespace, like *eg:prop1* and *eg:class2*. Given the relabelled model, we tasked ChatGPT:

> **Prompt 3:** Given the RDF/Turtle model below, please replace all properties and classes with the most likely standard ones. <rdf-model>

With ChatGPT-3 only 2/5 iterations succeeded in carrying out all substitutions. In those succeeding cases, the quality was still not as expected because of limited ontology reuse: Only IRIs in the example namespace were introduced, rather than reusing the *foaf*, *vcard*, and *org* vocabularies. Yet, the ad-hoc properties and classes were reasonably named, such as *eg:firstName*, *eg:countryName* or *eg:departmentName*. In contrast, ChatGPT-4 delivered better results: All classes and properties were substituted with those from standard vocabularies—foaf, vcard, and org were correctly identified. For some iterations, ChatGPT-4 used the schema.org vocabulary instead of the org vocabulary as an alternative approach.

4.2 Token Counts for Knowledge Graphs Schemas

After the results with the small custom knowledge graph we wanted to check the size of some well known knowledge graphs with respect to LLMs.

The LLMs behind ChatGPT can handle at the moment only 4096 tokens (GPT-3.5[3]) or 8192 respective 32,768 tokens for GPT-4[4].

Table 2. Token counts for selected knowledge graphs and serialisations

Graph	Serialisation Type	Token Count
Mondial Oracle DB schema	SQL schema	2,608 token
Mondial RDF schema	turtle	5,339 token
Mondial RDF schema	functional syntax	9,696 token
Mondial RDF schema	manchester syntax	11,336 token
Mondial RDF schema	xml/rdf	17,179 token
Mondial RDF schema	json-ld	47,229 token
Wine Ontology	turtle	13,591 token
Wine Ontology	xml/rdf	24,217 token
Pizza Ontology	turtle	5.431 token
Pizza Ontology	xml/rdf	35,331 token
DBpedia RDF schema	turtle	471,251 token
DBpedia RDF schema	xml/rdf	2,338,484 token

We counted tokens for various public knowledge graphs in different serialization formats with the library *tiktoken*[5] as recommended for ChatGPT. Tab. 2

[5] https://github.com/openai/tiktoken

lists the token counts for a couple of combinations ordered by token count. More data and information is available in the Supplemental Online Resources. The turtle serialization seem to result in minimal token count, but is still bigger than the similar SQL schema added for comparison. All knowledge graphs exceed the token limit for GPT-3.5 and 3 of 4 knowledge graphs listed here exceed the limit for GPT-4.

4.3 SPARQL Query Generation for the Mondial Knowledge Graph

In addition to the experiments with the small custom knowledge graph (see Sect. 4.1) we tested ChatGPT with the bigger mondial knowledge graph[6] which is published since decades with the latest "main revision" 2015.

We asked ChatGPT to generate a SPARQL query for a natural language question from a sparql university lecture[7]. We used the following prompt five times with ChatGPT-3 and ChatGPT-4 each:

> **Prompt 4:** Please create a sparql query based on the mondial knowledge graph for the following question: which river has the most riparian states?

The results are documented in the Supplemental Online Resources together with detailed comments on the given queries. Tab. 3 gives some statistics. In summary, all SPARQL queries given by ChatGPT were syntactically correct, but none of them worked when executed. Actually all queries had at least one error preventing the correct execution like referencing a wrong namespace, wrong usage of properties or referencing undefined classes.

Table 3. Findings in generated sparql queries for prompt 4.

	ChatGPT-3	ChatGPT-4
syntactically correct	5/5	5/5
plausible query structure	2/5	4/5
producing correct result	0/5	0/5
using only defined classes and properties	1/5	3/5
correct usage of classes and properties	0/5	3/5
correct prefix for mondial graph	0/5	1/5

[6] https://www.dbis.informatik.uni-goettingen.de/Mondial
[7] https://www.dbis.informatik.uni-goettingen.de/Teaching/SWPr-SS20/swpr-1.pdf

4.4 Knowledge Extraction from Fact Sheets

As an experiment to evaluate knowledge extraction capabilities, we used PDF fact sheets of 3D printer specifications from different additive manufacturing (AM) vendor websites. The goal is to build a KG about existing 3D printers and their type as well as capabilities. We fed plaintext excerpts (extracted via pdfplumber) from these PDFs into ChatGPT-3 and prompted it to:

> **Prompt 5:** Convert the following $$vendor$$ 3d printer specification into a JSON_LD formatted Knowledge Graph. The node for this KG should be Printer as a main node, Type of 3d printer such as FDM, SLA, and SLS, Manufacturer, Material, Applications, and Technique.

Since the fact sheets are usually formatted using a table scheme, the nature of these plain texts is that mostly the printer entity is mentioned in the beginning of the text which then is further characterized in a key-value style. As a result, the text typically does not use full sentences and contains only one entity that is described in detail, but several dependant entities (like printing materials). However, the format of the key-value pairs can be noisy. Key names can be separated with colons, new line feeds, or in contrast multiple key-value pairs can be in the same line, which could impose a challenge. Nevertheless, ChatGPT was able to identify the key-value pairs of the evaluation document in a reliably way. Unfortunately, it delivered out of 5 test runs for this document 4 partial and 1 complete JSON document. In spite of that, we summarize first insights gained from a knowledge engineering perspective (but for the sake of brevity, we refer to the output documents in the experiment supplements)

– The JSON-LD output format prioritizes usage of schema.org vocabulary in the 5 evaluation runs. This works good for well-known entities and properties (e.g. `Organization` @type for the manufacturer, or the `name` property), however, for the AM-specific feature key names or terms like `printer` ChatGPT-3 invents reasonable but non-existent property names (in the schema.org namespace) instead of accurately creating a new namespace or using a dedicated AM ontology for that purpose.
– Requesting `turtle` as output format instead, leads to different results. E.g. the property namespace prefix is based on the printer ID and therefore printer descriptions are not interoperable and can not be queried in unified way in a joint KG.
– Successfully splitting x, y and z values of the maximum print dimension (instead of extracting all dimensions into one string literal) works in 3 runs. Although ChatGPT-3 accurately appends the unit of measurement to all x, y, z values (which is only mentioned after the z value in the input) in those cases, this is a modelling flaw, as querying the KG will be more complex. In one run it addressed this issue by separating units into a separate unit code field.

– A similar effect was observed when it comes to modelling the dependent entities. E.g., in 4 runs, the manufacturer was modelled correctly as a separate typed entity, in 1 as string literal instead.

As a general conclusion of the experiment, ChatGPT-3 has overall solid skills to extract the key value pairs from the sheets, but the correct modelling or representation in terms of a KG significantly varies from run to run. Subsequently, none of the generated JSON documents contained sufficient information on their own, but only a subset that was modelled accurately. A question for future research is whether cherrypicking of individual JSON elements from outputs of several runs and combining them into one final document or iteratively refining the output by giving ChatGPT generic modelling feedback (like use an ontology, or separate unit information, etc.) can be automated in a good and scalable way.

4.5 Knowledge Graph Exploration

Experts in the field of knowledge graphs are familiar with concepts from RDF Schema (RDFS) (domain/range, subPropertyOf, subClassOf) and Web Ontology Language (OWL) (ObjectProperty, DatatypeProperty, FunctionalProperty, ...). Often, each of these experts has their preferred tools and methods for gaining an overview of an ontology they are not yet familiar with. We asked ChatGPT-3 two different questions requesting the mermaid[8] visualization of the most important concepts and their connections:

> **Prompt 6:** Can you create me a visualization showing the most important classes and concepts and how they are linked for dbpedia ontology, serialized for mermaid?

> **Prompt 7:** Can you create me a visualization of the most common concepts of the DBpedia ontology and their connections focusing on domain and range defined in properties.

We expected a graph with at least eight nodes and their corresponding edges. The identifiers for the nodes and edges are expected to follow the Turtle or SPARQL `prefix:concept` notation. If the first question did not achieve the goal, we asked additional questions or demands to ChatGPT-3. The results are presented in Tab. 4 and we evaluated the displayed graphs based on the following criteria:

[8] "...a JavaScript-based diagramming and charting tool ..." https://mermaid.js.org/.

Table 4. Diagram content overview.

	Prompt 6	Prompt 7
Mermaid Type	graph	graph*
Labels of Nodes	prefix and concept	prefix and concept**
Labels of Edges	prefix and concept	prefix and concept**
Number of Nodes (total/existing/dbo)	10/10/8	13/12/12
Number of Edges (total/unique)	12/2	17/17

* One more prompt was needed to serialize a graph
** One more prompt was needed to add prefixed labels

Prompt 6 led to an answer with a hierarchical graph representation of the important classes defined in the DBpedia ontology. The diagram already met our requirements regarding the minimum number and labelling after the first answer and can be seen in the Supplemental Online Resources.

The class hierarchy was represented by the `rdfs:subPropertyOf` relation, and the nodes were labelled in prefix notation, as were the edges. By arranging it as a tree using the subClassOf-pattern, only two different properties were used for the relations (edges). The root node was of type `owl:Thing` other nodes are connected as (sub)classes from the DBpedia ontology. These were: Place, Organization, Event, Work, Species, and Person. The class Work had one more subClassOf relation to the class MusicalWork. The class Person had the most complex representation, with two more subClassOf relations leading to `foaf:Person` and `foaf:Agent`, the latter of which is a subclass of the root node (`owl:Thing`).

In the second prompt (Prompt 7 ChatGPT-3 referred to a graphic file within the answer text that no longer existed. Upon further inquiry, a mermaid diagram was generated. It was of type "Graph" and contained thirteen common concepts and seventeen edges, which were all unique. The labels of both, nodes and edges contain no prefixes, but were addable with further inquiry. Only the generated concept `dbo:Occupation` is non-existent. All remaining nodes and edges comply with the rules of the ontology, even if the concepts used are derived through further subclass relationships. The resulting diagram is shown in the Supplemental Online Resources.

While prompt 6 leads to a result that can be more comprehensively achieved with conventional tools for visualizing RDF, the result from prompt 7 provides an overview of concepts (classes) and properties that can be used to relate instances of these classes to each other.

5 Conclusion and Future Work

From the perspective of a knowledge graph engineer, ChatGPT has demonstrated impressive capabilities. It successfully generated knowledge graphs from semi-structured textual data, translated natural language questions into syntactically correct and well-structured SPARQL queries for the given knowledge graphs, and even generated overview diagrams for large knowledge graph schemas, as outlined in Sect. 4. An detailed analysis revealed that the generated results contain mistakes, of which some are subtle. For some use cases, this might be harmless and can be tackled with additional validation steps in general, like with the metrics we used for SPARQL queries. In general, our conclusion is, that one needs to keep in mind ChatGPT's tendency to *hallucinate*[9], especially when applied to the field of knowledge graph engineering where many engineers are used to mathematical precision and logic.

The closed-source nature of ChatGPT challenges scientific research on it in two ways: 1. Detailed capability ratings of closed-source probabilistic models require much effort 2. Result reproducibility is bound to service availability and results might be irreproducible at a later date (due to service changes)

Thus, open training corpora and LLMS are mandatory for proper scientific research.

In the future, metrics are to be found to rate generated ChatGPT answers automatically, like we broached with SPARQL queries. This again enables to extend the number of test cases for a specific experiment and to generate profound statistical results. Another research focus should be given to methods that let the LLM access a broader/necessary context to increase the chance for correct answers.

Acknowledgements This work was partially supported by grants from the German Federal Ministry for Economic Affairs and Climate Action (BMWK) to the CoyPu project (01MK21007A) and KISS project (01MK22001A) as well as from the German Federal Ministry of Education and Research (BMBF) to the project StahlDigital (13XP5116B) and project KupferDigital (13XP5119F).

Supplemental Online Resources

Details on the experiments described can be found at the following github repo: https://github.com/AKSW/AI-Tomorrow-2023-KG-ChatGPT-Experiments

References

1. Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (Mar 2023). Sparks of artificial general intelligence: Early experiments with gpt-4. https://doi.org/10.48550/ARXIV.2303.12712.
2. Cagle, K. (Mar 2023). Nine chatgpt tricks for knowledge graph workers. https://www.thecaglereport.com/2023/03/16/nine-chatgpt-tricks-for-knowledge-graph-workers/, Accessed: 14. Apr. 2023.
3. Ekaputra, F. J., Llugiqi, M., Sabou, M., Ekelhart, A., Paulheim, H., Breit, A., ... & Auer, S. (Mar 2023). Describing and organizing semantic web and machine learning systems in the swemls-kg. https://doi.org/10.48550/ARXIV.2303.15113.
4. Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., & Ahmad, H. (Dec 2022). "i think this is the most disruptive technology": Exploring sentiments of chatgpt early adopters using twitter data. https://doi.org/10.48550/ARXIV.2212.05856.

[9] Generation of content without any foundation

5. Kaliciński, K. (Jan 2023). Create neo4j database model with chatgpt. https://neo4j.com/developer-blog/create-neo4j-database-model-with-chatgpt/, Accessed: 14. Apr. 2023.
6. Lin, W., Babyn, P., yan, Y., & Zhang, W. (Mar 2023). Context-based ontology modelling for database: Enabling chatgpt for semantic database management. https://doi.org/10.48550/ARXIV.2303.07351.
7. Omar, R., Mangukiya, O., Kalnis, P., Mansour, & E. (Feb 2023). Chatgpt versus traditional question answering for knowledge graphs: Current status and future directions towards knowledge graph chatbots. https://doi.org/10.48550/ARXIV.2302.06466.

Foundations for the Development of an AI-based, Platformindipendent cOmpanion-app [for] Lifelong Learning-Optimization (APOLLO)

Dominik B.O. Boesl[1(✉)], Teresa Achtenberg[2] and Lisa Bergler[1]

[1] Hochschule der Bayerischen Wirtschaft (HDBW) gGmbH, Konrad-Zuse-Platz 8, 81829 München, Deutschland
{dominik.boesl,lisa.bergler}@hdbw-hochschule.de
https://project-apollo.de/

[2] Innomancer GmbH, Olympiastr. 6a, 86179 Augsburg, Deutschland
teresa.achtenberg@innomancer.com

Abstract. In today's knowledge society, the necessity for lifelong learning can be taken for granted. Hence, everybody will be facing the challenge of managing, organizing an optimizing the individual lifelong learning journey. Apart from teaching the essential methodological skills and abilities to enable lifelong acquisition of knowledge and skills, more and more guidance is needed to get an overview of the overabundance of learning content from diverse suppliers.

This paper elaborates on the the basic assumptions, analyses, and framework conditions for the development of an "AI-based, Platformindipendent cOmpanion-app [for] Lifelong Learning-Optimization" (acronym: APOLLO) as part of a 36-month funded project with kind financial support from the German Federal Ministry of Education and Training (BMBF) under coordination by the German Federal Institute for Vocational Education and Training (BIBB) as part of the innovation framework program "INVITE". It outlines the project idea, describes the motivation and problem definition reflected in the educational domain as well as providing an overview of the current national and international state of research. All these aspects have been taken into consideration when defining the functional scope of the application.

Zusammenfassung. In der heutigen Wissensgesellschaft kann die Notwendigkeit zur Befähigung zum lebenslangen Lernen als gegeben angesehen werden. Abgesehen von der Vermittlung der essenziellen methodischen Fähig- und Fertigkeiten, um den lebenslangen Wissens- und Kenntniserwerb zu ermöglichen, ist aber auch immer mehr Lotsenhilfe beim Verschaffen eines Überblicks über das überbordende Angebot an Lerninhalten unterschiedlichster Anbieter nötig. Dieses Paper skizziert die grundlegenden Annahmen, Analysen und Rahmenbedingungen für die Entwicklung einer Ai-basierten, Plattformübergreifenden cOmpanian-app [für] Lebenslange

C. Zinke-Wehlmann und J. Friedrich (Hrsg.), *First Working Conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow*, AIDRST 2023, Informatik aktuell, S. 117–127, https://doi.org/10.1007/978-3-658-43705-3_9

Lern-Optimierung (Akronym: APOLLO) im Rahmen eines 36-monatigen Förderprojektes mit freundlicher Unterstützung des Bundesministeriums für Bildung und Forschung (BMBF) und unter Betreuung durch das Bundesinstitut für Berufliche Bildung (BIBB) als Teil der Förderrichtlinie des Innovationswettbewerbs INVITE. Es umreißt die Projektidee, schildert die Motivation und Problemstellung im Bildungsbereich sowie den aktuellen nationalen und internationalen Forschungsstand, die zur detaillierten Spezifikation des Funktionsumgangs herangezogen wurde.

Keywords: lifelong learning · AI-assisted learning · skill-assessment · Artificial Intelligence · learning paths · individual learning path recommendation · ESCO · skill profile · taxonomies · recommender systems

1 Brief Description of the Project Idea

The APOLLO app undesrstands itself as a "digital travel companion for the lifelong, individual learning journey". The goal of the project is therefore to provide people with an AI-based, intelligent assistant that accompanies them on their entire, lifelong education and training path. An individualized, representative skills profile created on the basis of data collected over the long term reflects the personal competencies and development potential of the users and supports them in career planning and development. Lifelong learning thus becomes transparent, plannable and manageable; users are empowered to use their maximum personal potential for professional (personal) development, to consciously plan and actively shape their career path in the medium and long term. This enables them to react early and proactively to changes in their own life situation or unforeseeable socio-economic changes (e.g. digitalization, COVID, ...). Apart from the personal AI assistant to be developed, they will be supported by a variety of services, such as consulting offers or further training suggestions. The human being with his or her individual abilities and skills, hence, returns to the focus of the world of employment, without losing the economic interests of employers out of sight.

A consortium consisting of five partner organizations is working on the development of the cross-platform APOLLO-app (iOS, Androing and possibly a web interface will be supported). The consortium lead by the Hochschule der Bayerischen Wirtschaft (HDBW) consists of the Bildungswerk der Bayerischen Wirtschaft (bbw), the Bildungswerk der Baden-Württembergischen Wirtschaft (BiWe), the TÜV Rheinland Akademie and the Bertelsmann Stiftung. These five organizations are combining their expertise in the educational domain, the data-treasure of hundreds of thousands of data sets as well as the already existing results of other recent (research) projects to create a personalized education guide of the future for everybody's pocket.

2 Motivation and Problem Definition

Lifelong learning is becoming increasingly important in people's lives. For several decades now, it has been apparent that the required flexibility in everyday working life as well as the ability to reorient and adapt to changing working conditions with newly emerging skill-profiles are becoming increasingly important [1]. This development has been accelerated by the rapid innovations in the field of digitalization [2].

Some researchers even see the (today so common) fixed company affiliation – at least in the segment of top performers – disappearing and predict a shift to cross-company, project-based collaboration in loose organizational or merely contractual structures; permanent employees will work together with hired or leased external experts and other bought workforce to solve problems. In this agile, volatile and short-lived labor market of the future, however, only those who are willing and able to improve themselves further will be able to survive [3]. Only employees constantly acquiring and updating the necessary skills and capabilities for the respective work environment will be able to keep up and succeed [4].

In the area of the so-called 'Individual Contributors (ICCs)' (highly qualified, multi-nationally deployable and globally sought-after top talents), the challenge for those concerned will be limited – in this domain, the competition for the best experts has already flared up and is commonplace. Fueled by the digital transformation, which is bringing about fundamental changes in for many occupational profiles or even their complete substitution, adaptability, flexibility and the will for further development will become important competitive factors on the labor market in the future [2]. In the short term, this trend will propagate throughout the entire qualification pyramid, and thus also in medium- to low-skilled occupational fields. A successfully completed vocational training can and will soon no longer provide a sufficient guarantee to remain employable in the long run, but still represents the necessary basis for an own gainful employment [5]. Therefore, the overriding goal should be to acquire a degree and to keep it up to date. Only those who actively continue their education, adapt to new situations and continuously develop their skill profile will be able to survive on this labor market in the long run [6].

Lifelong learning is the consequence. However, not all employees and job seekers have internalized this necessity, yet. APOLLO provides approaches to address the basic challenges in this context in a timely, efficient and motivating way. The target group is to be sensitized for the necessity of lifelong (further) education, motivated to interact, presented with the appropriate information and individualized content, and the progress moderated and guided accordingly. The employee regains the power to act (instead of reacting) – accompanied by a personal, artificial and intelligent training assistant, the interest of the individual takes center stage and the focus is on securing or optimizing his employability – in accordance with individually set career goals [7]. Based on a lifelong growing skill profile, this assistant accompanies the employee, advises and coaches; manages and protects personal data; and is a navigation aid in the jungle of Germany-wide further education offers.

2.1 Starting Point: Learning and Adaption are Life Tasks

Even a completed vocational training and what has been learned through it is no longer sufficient to meet the requirements of professional life in the long run. Employability throughout the entire working life can only be achieved if everyone continuously develops his or her competences according to the changing challenges [8]. It should be noted that this personal development is not only explicit (in the form of trainings etc.), but also implicit (during the employment phase, daily and 'on-the-job'). The understanding of the importance and necessity of lifelong learning must be anchored early in education and, hence, become a fixed, integral part of school curricula. It must be part of the basic educational mission of every school to impart the necessary competencies and skills for continuous, self-motivated further education. This is the only way to ensure that the imbalance on the labor market is not exacerbated by personal aptitude and access to further education (e.g. restricted by social status). Conversely, the comprehensive implementation of appropriate measures can simplify access to education and training and thus actively contribute to a democratization of the labor market.

2.2 Initial Situation: The Educational Market is Very Differentiated

According to the German Institute for Adult Education (Deutsches Institut für Erwachsenenbildung), there are approximately 25,000 continuing education providers. The spectrum ranges from non-profit to commercial, from regional to nationwide and international, from micro enterprises to large institutions [9]. This is accompanied by a considerable variety of types of offers, certificates and degrees as well as different terminology. In principle, this differentiation is a great strength, since it allows a very individual response to different demand situations. However, it must be ensured that this individual need situation and the respective suitable offer also match. There are many regional and supra-regional as well as target-group-specific platforms that are intended to remedy this situation [10]. However, they have not yet succeeded in matching needs and qualification offers to satisfactory extent.

2.3 Target Group Analysis: Barriers to Participation in Continuing Education

In the context of the target group analysis, various personas including their customer journeys were created on the basis of the data provided internally using methods from design thinking. This resulted in various person-specific criteria and characteristics which were taken into account in the target group analysis. In the following, the central aspects of these analyses are briefly presented.

When it comes to increasing participation in continuing education, it is important to consider the individual barriers to participation in continuing education. There are many aspects to this. Examples include, that the person is not aware of a concrete need for further education or that the existing learning concepts offered do not seem to fit. Time constraints due to family and/or work commitments are obstacles as well as offers that are perceived as too expensive and unfulfilled counseling wishes. If, in addition, there is no support (financial or otherwise) from the employer or the employment agency, the person often fails to identify a suitable educational offer.

2.4 Target Group Analysis: Motivation for Participation in Continuing Education

The individual benefit is decisive for the participation in further training. Therefore, the three most important motives are maintenance qualification (further training to maintain the ability to work in one's own field of activity), adaptation qualification (further training to be able to meet changed or changing requirements of one's own work) and promotion qualification (further training to take over other/higher-value tasks) [8]. The aim of the APOLLO project is to encourage the target group to take responsibility for their own employability and professional development and for lifelong learning. Often, the honest and neutral answering of these questions is hampered by a divergence in the self-perception and the perception by others of one's own competencies. A neutral assessment of the users' abilities and skills is – even with the help of selective placement tests or other assessment procedures – usually only very inaccurate and therefore not satisfactory. This is where APOLLO comes in. The companion app suggests suitable training opportunities based on the user's individual skill profile. This AI-based profile, which is generated from a large number of data points, accompanies the person throughout his or her entire lifelong education and thus provides a realistic basis for matching education and job offers with the individual's qualifications.

However, end-users are not the only stakeholders who benefit from a networked, intelligent, AI- and Big Data-driven education solution. Educational institutions can use it to continuously optimize their curricula on the basis of the resulting competency profiles and constant monitoring of learning progress and objectives, and adapt them dynamically and flexibly to the needs of learners as well as companies and supra-regional sponsors (employment agencies, etc.). In addition, the transparent recording of potential demand situations can lead to forward-looking, predictive capacity planning at educational institutions. For companies – regardless of their size – the use of the infrastructure makes sense on several levels: on the one hand, they can offer incentives for those interested in further education to develop into the respective professional field by indicating personnel requirements; possibly, partnerships (similar or the same as the dual study) can even be concluded at an early stage. In the case of applications, neutral, fact-based analyses of the suitability of candidates can be carried out on the basis of their skill profiles and strengths and development potentials can be identified, which can then be included in a career development plan at an early stage. The skill profiles can also provide a neutral basis for transparent, fair, performance-related remuneration and salary components.

2.5 Methodological Approach and Data Basis

For the development of the APOLLO application solution, a modern, agile, SCRUM-based development process is followed. The development activity is divided into several phases, each of which focuses on one aspect of the APOLLO solution. In line with the agile development philosophy andm the design thinking methodology, this phase follows a user-centric paradigm: For APOLLO, the maximization of the user benefit is always predominant. This is illustrated, among other things, by the modeling of a user journey, which – based on a persona analysis – captures the requirements, desires and user processes of the individual stakeholders. From the beginning, close cooperation with the first pilot customers was essential to identify all diverse requirements as early as possible. It is tried to ensure the basic functionality of the application in a conventional, proven way according to the established principles of computer science. In later iterations, smart experimental approaches based on artificial intelligence and machine learning will be successively added.

For the modeling and mapping of the skill profiles, the consortium can rely on the expertise and work of the Bertelsmann Foundation. Therefore, the implementation will be done in close coordination and under consideration of the ESCO framework as well as the ONET skill taxonomies. So far, the mapping of the skill profiles in a graph database is planned. User skills and abilities are also stored in the knowledgebase and enriched with the corresponding ontologies according to the ESCO definition. From the attached architecture and concept visualizations the preliminarily assumed structure and the corresponding data flow can be taken: The user's skill profile is created (as automatically as possible) from different sources (credentials, certificates, self- and third-party assessments, ...). One challenge will certainly be in the recording of non-standardized competencies. (e.g., job references and letters of recommendation in prose), since the contents of these must first be extracted, analyzed, and classified by complex, AI-based methods (e.g., natural language processing). An elementary part of the data structure is therefore the representation of the goals as well as the possibility to categorize the user by his parameters (skill profile, goals, etc.) and to derive, monitor and control his learning journey based on this data. The required data is used to give a weighting to the entities (skills) or nodes. The requirements foresee the validation of credentials and/or certificates via blockchain by the education providers. These also receive a weighting and are reliable data points, especially when comparing and applying analytical procedures from statistics. From the weighted references of the knowledgebase, the digital image (quasi a digital twin) of the user or his competencies, referred to as user knowledge, is created. This represents the core of the user profile.

3 Analysis of the National and Intern. State of Research

The analysis of the national and international state of research leads to a mixed picture. In the business environment, the use of skill profiles has become widespread, especially in larger companies. They are usually used as a proven tool in the context of the annual performance evaluation (also as a basis for the determination of personal goal achievement), for the identification of development potential and as a basis for the trans- parent description and filling of vacant positions in the company [11]. However, the tools currently in use (e.g., from SAP, Persis, or Evidenz) only rely on manually created and maintained skill profiles, and the assignment of employees to career paths, the monitoring of individual development perspectives, and the implementation of training and development measures are also performed by the respective HR consultant [12]. AI support, the use of big data analytics or process automation is – at best – rudimentarily indicated but by no means widely available. There is also no networking with cross-company or external training offers and platforms. In the educational environment (e.g. at colleges and unive)rsites, however, skill profiles are almost unknown, especially in the German-speaking area – apart from the rudimentary use in the allocation of university places for courses of study with restricted admission. Especially in view of the aforementioned widespread use in the professional environment, this is alarming, since the participants in the educational system can neither benefit from the use in the context of per- sonalized education and training nor are they prepared for the use in the company. Nor is there any sensitization about the necessity of or preparation for later lifelong learning [13].

Only in the university domain of (automated) study support / student counselling simple standard questions ('Are you more interested in economics or in natural sciences?') are used in isolated systems, although corresponding solution approaches, e.g. for the automation of first- and second-level support in customer hotlines, have already been tested millions of times. The fact that these problems are often assigned as selective bachelor and master theses, later piloted as rudimentary solutions and not professionally implemented as part of an overall concept, certainly does not contribute to an increase in quality.

Whereas, the necessary basic technologies are available, already tested, robust and suitable for productive use – but there is a lack of definition of an adequate overall package taking into account the requirements and including the necessary interdisciplinary , domain-specific knowledge about the education and training area.

In international comparison, there exist – in the university environment – a manageable number of positive case studies that address some of these aspects: At Georgia Tech University in the USA, students are already successfully supported and taught by 'Jill', a virtual tutor [14]. The quality of the virtual learning environment is so good that students can sometimes no longer tell the difference between it and conventional lecturer support [14]. Genie', a virtual companion at Deakin University in Australia, even goes one step further and offers users interactive help in designing curicula, selecting additional modules and individualized, interest-based scheduling of intra- and extrauniversity activities [15],[16] [17]. Proven, existing approaches from the industrial environment can be transferred to the entire lifelong training path in a goal-oriented and methodologically-didactically sound way. This promises better individual support, transparent insights into and full control over their competence profile as well as the resulting possibility to create personalized learning paths and training roadmaps.

4 Description od Specific App Idea

The aim of the APOLLO project is to accompany people on their education and training path and to support them in the selection of individually tailored training measures as well as in the organization of their lifelong learning. The individual and his or her needs are at the center of our efforts. People interested in further education should be enabled to regain control over their personal development, thereby strengthened in their role as employees and enabled to reflect and develop their personal strengths and interests as fully as possible in a professional profile. It is paramount that the user always has full control and decision-making power over his data and its use. In case of doubt, all interactive components must always act in the sense of data protection and for (instead of against) the user. But a win-win-win situation must also be created for the employing companies as well as for the providers and promoters of vocational and further training. The project is therefore divided into three main components.

The definition and development of individualized skill profiles – AI-supported and with the help of available data (Big-Data) – for the transparent mapping of the abilities and skills of the respective user. They serve, among other things, as a basis for the proposal of suitable further training measures in order to achieve the goals set by the user or as support for the assessment of the personal suitability for the application for advertised jobs. The assistant (which is available to the users e.g. on mobile devices in the form of an app) accompanies the individual on his lifelong educational journey and represents the personification of the user interface. Depending on the preference, the Learning and Skill Companion (LSC) offers a target-group-appropriate approach (among other things with the help of the methods of serious gamification) and helps in the search and selection of further education offers, motivates in the implementation and helps to keep an eye on one's own progress.

The collected data holds great potential for the participating employers, the providers of training and further education as well as the supporting and supporting authorities. Focal points of interest of those interested in further education as well as the needs of the companies can be identified and responded to quickly. Similarly, companies can obtain a better overview of the personnel resources available on the labor market and their skill profiles; based on this, forward-looking personnel planning with corresponding roadmaps can be defined, bottlenecks can be better avoided and, ideally, investments can be made in corresponding personnel development or further training programs and partnerships at an early stage. Both variants represent a decisive added value for the employer side, because in many cases the companies do not have an adequate overview of the competence distribution of their own employees. Conversely, emerging changes in requirement profiles and job profiles can also be taken into account in the planning of curricula and training contents at an early stage, so that the (further) training managers can also act predictively and with foresight, plan capacities and adapt contents.

5 Conclusion and Outlook

Lifelong learning is a cornerstone of individual and personal development. In an increasingly fast-paced world, where innovation cycles shorten and technological progress sparks disruptive innovations manifesting on Megatrend-scale, humanity is affected globally and on all different levels of society [18]. When technologies like the evolution of mainstream information technology, the Internet, the (industrial) metaverse, robotics, automation, big data and artificial intelligence in all its forms of manifestation start permeating the human living realm, the implicitly affect the workspace. Skill-profiles are altered; new skills are required – sometimes even over night – and affect the whole workforce, often in very unexpected ways [19]. This does not solely apply to technological disruption, though it is usually the fastest, most agile driver of changing skill requirements on the labor market. But also other Megatrends, that affect societies on socio-political, socio-cultural, ethical-moral and socio-economic level, carry these properties. They can act as accelerators as well as inhibitors for societal change [20] [21] – but nevertheless this is what they bring: transformation.

Motivated by these fundamental insights, important interim results have already been achieved within the past 20 months of development, including successful alpha testing as part of the 'BIBB INVITE Tool Check'. With the help of several development iterations, it was possible to develop a successfully running prototype, which is still being improved and adapted. The basic assumptions presented in this paper can still be regarded as valid, and the various projects within the INVITE project have produced similar interim results. Due to the necessity of lifelong learning, there are still relevant questions for the APOLLO project. Therefore, it is necessary to generate the relevant data with the help of the research project and to put them into context.

However, some challenges still lie in the technical depth: The definition of suitable, accurate skill profiles is not only a concern of the consortium publishing here. The disruptive developments in the field of AI algorithms in recent weeks and months – especially in the area of transformer-based language models and the associated analysis and generation capabilities in natural language interactions – shed new light on the technological core of proposal generation. These aspects are the subject of ongoing, continued research in the consortium and will be addressed within further publications and papers.

Acknowledgments. Project APOLLO is kindly funded by the German Federal Ministry of Education and Research (BMBF) as part of the innovation framework "INVITE" which is coordinated by the Federal Institute for Vocational Education and Training (BIBB).

References

1. Billett, S. (2010). The perils of confusing lifelong learning with lifelong education. *International Journal of Lifelong Education, 29*(4), 401–413. https://doi.org/10.1080/02601370.2010.488803
2. Boesl, D. B. O., & Liepert, B. (2016). 4 Robotic Revolutions – proposing a holistic phase model describing future disruptions in the evolution of robotics and automation and the rise of a new Generation 'R' of Robotic Natives. *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Daejeon. https://doi.org/10.1109/iros.2016.7759209.
3. Boesl, D. B .O., Bode, M., Liepert, B., & Greisel, S. (2017). Structured Megatrends Research as Foundation for Future-Oriented Research-Planning and R&D Roadmapping in Robotics. *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*. Mysore. https://doi.org/10.1109/CTCEEC.2017.8455126.
4. Boesl, D. B. O., Bode, M., & Greisel, S. (2018). Drafting a robot manifesto – new insights from the robotics community gathered at the european robotics forum 2018. *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. Nanjing. https://doi.org/10.1109/ROMAN.2018.8525699.
5. Majhanovich, S., & Brook Napier, D. (2014). Lifelong learning in the new millennium: voices from the XV World Congress of Comparative Education Societies. *International Review of Education, 60*, 449–462.
6. Steffens, K. (2015). Competences, learning theories and MOOCs: Recent developments in lifelong learning. *European Journal of Education, 50*(1), 41–59.
7. Adiguzel, T., Kaya, H., & Cansu, F. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology* (15. ep429). https://doi.org/10.30935/cedtech/13152.
8. Jarvis, P. (2010). *Adult Education and Lifelong Learning: Theory and Practice* (4th ed.), Routledge.
9. Wie viele Weiterbildungsanbieter gibt es eigentlich in Deutschland? Daniela Jung Presse- und Öffentlichkeitsarbeit, Deutsches Institut für Erwachsenenbildung vom 27.06.2008, https://idw-online.de/de/news267838. Aufgerufen am 11.04.2023.
10. Pottmer, B., Antony, P., Beyer, M., Brümmer, U., Heister, M., Kruppe, T., & Schroeder, W. (2019). Weiterbildung 4.0 Solidarische Lösungen für das lebenslange Lernen im digitalen Zeitalter. *böll.brief TEILHABEGESELLSCHAFT,8*, S. 18 ff.
11. Webquelle: Europass Profile. https://europa.eu/europass/en/europass-tools/europass-profile, abgerufen: 14.06.2023, 22:34:27 CET.
12. Boyatzis, R. E., & Kolb, D. A. (1995). From learning styles to learning skills: The executive skills profile. *Journal of managerial psychology, 10*(5), 3–17.
13. Boyatzis, R. E., & Kolb, D. A. (1991). Assessing individuality in learning: The learning skills profile. *Educational Psychology, 11*(3–4), 279–295.
14. Goel, A. K., & Polepeddi, L. (2018). Jill Watson: A virtual teaching assistant for online education. In *Learning Engineering for Online Education* (pp. 120–143). Routledge.

15. Webquelle: EMPRIZE at Georgia Tech University. https://emprize.gatech.edu, abgerufen: 17.06.2023, 09:52:19 CET.
16. Webquelle: Online Learning Tools at Georgia Tech University. https://dilab.gatech.edu/a-suite-of-online-learning-tools/, abgerufen: 15.06.2023, 19:17:48 CET.
17. Bonfield, C. A., Salter, M., Longmuir, A., Benson, M., & Adachi, C. (2020). Transformation or evolution?: Education 4.0, teaching and learning in the digital age. *Higher Education Pedagogies, 5*(1), 223–246.
18. Boesl, D. B. O., & Bode, B. A. M. (2016). Generation R: Why our grandchildren will grow up as the first Generation of Robotic Natives. *2016 IEEE International Conference on Emerging Technologies and Innovative Business Practices for the Transformation of Societies (EmergiTech)* (417–420). Balaclava, Mauritius. https://doi.org/10.1109/EmergiTech.2016.7737377.
19. Móga, K. (2021). Augmented/Mixed Reality Technologies Supporting Digital Surgery. *2021 IEEE 19th International Symposium on Intelligent Systems and Informatics (SISY)* (pp. 183–189). Subotica, Serbia. https://doi.org/10.1109/SISY52375.2021.9582533.
20. Boesl, D. B. O., & Haidegger, T. (2021). Automating the Achievement of SDGs: Robotics Enabling and Inhibiting the Accomplishment of the SDGs. In *Emerging Science, Frontier Technologies, and the SDGs – Perspectives from UN System and Science and Technology Communities* (pp. 122–126). The UN Interagency Task Team on Science, Technology and Innovation for the SDGs (IATT), New York (NY).
21. Mai, V. (2022). The Role of Robotics in Achieving the United Nations Sustainable Development Goals-The Experts' Meeting at the 2021 IEEE/RSJ IROS Workshop [Industry Activities]. *IEEE Robotics and Automation Magazine, 29*(1), 92–107. https://doi.org/10.1109/MRA.2022.3143409

Viability of Knowledge Management Practices for a Successful Digital Transformation in Small- and Medium- Sized Enterprises

Oluwatomiwa Oni-Orisan(✉)

Leipzig University, 04109 Leipzig, Germany
Oni-Orisan@wifa.uni-leipzig.de

Abstract. Digital innovations and technologies, particularly artificial intelligence, offer a unique opportunity to fundamentally transform business processes. Small and medium-sized enterprises (SMEs), having a significant impact on the German economy, are encouraged to fully embrace this opportunity for digital transformation.

Inspired by the usage of knowledge management as a mediation mechanism for effective AI application in [5], this case study examines its practical implications. Wiewald, an SME specializing in compressor systems, serves as an application partner in the KMI project, exploring the implementation of AI in SMEs.

By comparing academic literature on Industry 4.0's impact on knowledge management with industry experts' perspectives, we develop an appropriate digitalization strategy based on knowledge management. Its potential implementation is discussed using Wiewald as a practical example.

Zusammenfassung. Diese umfassende Fallstudie taucht in das Gebiet des Wissensmanagements (KM) ein und dessen Anwendung bei Wiewald, einem Unternehmen, das sich auf die Planung und Gestaltung von Kompressorsystemen spezialisiert hat. Die Studie präsentiert eine umfassende Untersuchung verschiedener Dimensionen des KM und bietet strategische Empfehlungen, die auf Wiewald's Digitalisierungsstrategie zugeschnitten sind.

Die Studie hebt die Bedeutung des KM im digitalen Zeitalter hervor, in dem die transformative Kraft der digitalen Revolution einen Paradigmenwechsel in der Fertigungsbranche herbeigeführt hat. Das Aufkommen des Internet-of-Things und der Zustrom von Massendaten haben den Übergang von Big-Data zu Smart-Data erforderlich gemacht, welche handlungsorientierten und aussagekräftigen Informationen repräsentiert. Es wird jedoch darauf hingewiesen, dass Organisationen das volle Potenzial von Smart Data möglicherweise noch nicht vollständig ausgeschöpft haben.

Die Fallstudie zieht Erkenntnisse aus empirischer Forschung mit malaysischen Fertigungsunternehmen heran und hebt die entscheidende Rolle der Integration des KM in Geschäftsprozessen hervor. Vier Dimensionen des KM werden identifiziert: Wissensschaffung und -erwerb, Wissensaustausch und -transfer, Wissensspeicherung und -ab- frage sowie Wissensanwendung. Diese Dimensionen sind miteinander verbunden und betonen die Bedeutung der Nutzung organisatorischen Wissens für effektive Innovationen.

C. Zinke-Wehlmann und J. Friedrich (Hrsg.): *First Working Conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow*, AIDRST 2023, Informatik aktuell, S. 129–139, 2024. https://doi.org/10.1007/978-3-658-43705-3_10

Die akademische Literatur zum KM wird durch eine bibliometrische Analyse erforscht, um Lücken zu identifizieren und aktuelle Trends sowie aufkommende Techno- logien zu integrieren. Die Studie betont die Notwendigkeit, die Kluft zwischen praktischen Leitlinien für die Wissensanwendung in der Industrie und dem Stand des Wissensmanagements in der akademischen Forschung zu überbrücken.

In Zusammenarbeit mit Wiewald konzentriert sich die Studie auf die Entwicklung eines Kompressorsystem-Konfigurators, der dem Cheftechniker bei der Planung und Gestaltung neuer Systeme helfen soll. Durch Machbarkeitstests unter Verwendung historischer Daten wurden in der Vergangenheit Entscheidungsbaummodelle entwickelt, die basierend auf grundlegenden Anforderungen wahrscheinliche Entscheidungswege bestimmen können. Dieser Konfigurator zielt darauf ab, Vorschläge für gültige Systeme zu liefern und schnelle Reaktionen sowie hochwertige Entscheidungsfindung zu ermöglichen.

Für die Implementierung des Konfigurators wird ein Ansatz namens „Actionable Cognitive Twin“ (ACT) vorgeschlagen, der die Transparenz von Informationen, die Entscheidungsunterstützung und die Agilität verbessern soll, während er gleichzeitig die Beteiligung der Mitarbeiter an der Gestaltung von Geschäftsprozessen und die Förderung von Innovationen anregt. Die Implementierung von ACT umfasst die systematische Verwaltung von Wissen, die Integration von Ontologien und Wissensgraphen sowie den Einsatz datengetriebener Techniken wie Natural Language Processing und neuronalen Netzwerken.

Keywords: Knowledge Management · Digital Transformation · Knowledge Application

1 Background

Digital innovations and technologies, especially artificial intelligence, present a unique opportunity to fundamentally change how organizations work: improving knowledge access, facilitating more extensive product and service development, minimizing bureaucratic obstacles and enabling the emergence of new business domains. Put simply, artificial intelligence plays a key role in facilitating a digital revolution towards a knowledge-driven economy.

Small- and medium-sized enterprises (SMEs) play a vital role in the German economy, accounting for 99.4 percent of the market and driving national competence development [1, 2]. Fully capitalizing on this opportunity for digital transformation is crucial given their significant influence, which is why the German government's strategic initiative for Industry 4.0 [4], aims to leverage and maximize SMEs' transformative potential. It is reasonable to assume that by embracing Industry 4.0, SMEs can ensure long-term competitiveness in the evolving business landscape, positioning themselves for sustained success.

Unfortunately [3], "Let There Be Change", sheds light on the limited AI implementation among German companies. The study reveals that a mere 12 percent of these companies consider their AI maturity level sufficient to improve their processes and organizations. In contrast, a substantial 63 percent remain in the experimental phase of adopting AI at an enterprise level, indicating their challenges in leveraging AI effectively.

The empirical study by [5] with 120 senior executives from Italian manufacturing firms highlights the crucial role of knowledge management practices (KMPs) in achieving AI maturity and leveraging its potential. The findings show that AI deployment positively influences KMPs, stimulating their higher levels while enhancing supply chain resilience (SCR) and manufacturing firm performance (MFP). This emphasizes the need for an effective digitalization strategy that prioritizes KMPs as a mediating mechanism. By leveraging KMPs to transform data into valuable knowledge, organizations can enhance decision-making and improve MFP and SCR outcomes.

The KMI project ("Künstliche Menschlich Intelligent" – federal grant no. 02L 19C500) champions sustainable AI; prioritizing worker benefits over profit-driven approaches like Industry 4.0. By augmenting worker capabilities, reducing menial tasks, and ensuring usability, sustainable AI aims to preserve employability, fostering a more worker-centred "humane" approach within the Industry 5.0 context, where worker benefits are valued alongside company profits [4].

Building upon the insights of [5], we explore the potential applicability of their findings in a real-life SME. Wiewald, a specialist in configuring, installing, and maintaining compressor systems, serves as our focus for this investigation. As an active application partner in the KMI project, our research revolves around enabling the sustainable integration of AI within SMEs, drawing from our own insights and experiences gained through our work with Wiewald, which have shed light on additional challenges commonly faced by SMEs in the project.

One such challenge is an insufficient 'data-mindedness', referring to the unawareness among SMEs about the opportunity to generate additional value from their data for various reasons. In the case of Wiewald, this could be attributed to unstructured data or the absence of digitization. Another challenge is the presumed absence of a structured and integrated process to externalize and document employee knowledge, particularly when tacit expert knowledge is required for projects involving new innovations such as AI.

Concerning these two issues, our experience at Wiewald and a new understanding concerning the importance of knowledge management practices for digital transformation, we believe that companies like Wiewald, need to rely on knowledge and its exploitation, to sustain a long-term advantage. While extensive knowledge management practices may not be essential for day-to-day operations, SMEs must proactively prepare themselves with a digitalization strategy to remain relevant in the evolving landscape of the German economy's digital revolution. This strategy ensures their long- term success in an emerging knowledge-based economy, where the ability to leverage knowledge and adapt to digital advancements becomes a crucial competitive advantage.

2 Hypothesis and Research Question

We anticipate that an improved availability and accessibility of data and information would concurrently improve the potential for value generation and MFP of SMEs, thereby creating the necessary foundation for digital transformation strategies. Hence, we hypothesize that a viable solution must include:

- A means for systematic knowledge management to capture explicit knowledge and preserve tacit expert knowledge within the company, explicitly integrating it into a business process model.
- A holistic measure for innovation management that encourages active involvement from workers, especially when structuring or modelling business data and business processes, ideally facilitating worker-driven initiatives for innovation management.

By reconsidering our approach regarding a successful digital transformation strategy in SMEs, an argument concerning the merit and strategic importance of knowledge management practices for the sustainable implementation of AI should be raised and evaluated, leading to the following research question:

"To what extent can the effective and decisive application of knowledge raise an SME's AI maturity level and determine the success of its digital transformation?"

Evaluating this research question involves assessing the impact of effective knowledge application on AI maturity and its correlation with the success of digital transformation efforts in SMEs. The examination will consider factors such as the percentage of explicitly captured business process knowledge compared to processes lacking explicit detailed information, as well as tracking the number of adjustments and innovations introduced by workers since the inception of the digitalization solution, along with their impact on MFP.

This research question delves into the relationship between knowledge management practices (KMPs), AI maturity, and the success of digital transformations in SMEs. It seeks to demonstrate the crucial role of KMPs in elevating an SME's AI maturity level and achieving a successful digital transformation.

To address this research question, the case-study will provide a concise and comprehensive overview of knowledge management, exploring its various dimensions, inter- connections, and their impact on driving process and product innovation. By doing so, it will examine how effective knowledge management enhances AI maturity within organizations and facilitates the development of impactful digital transformation strategies.

3 Theory: Global Perspective – Comparing Industry Experience with Structure of Academic Research

The digital revolution, driven by advancements in technology such as the Internet of Things, has brought about a transformative shift in the manufacturing domain [7]. It has enabled seamless communication and autonomous collaboration among machines, leading to a massive influx of data commonly known as big data. However, it's important to note that big data is not synonymous with smart data, which represents actionable and meaningful information. Despite the prolonged existence of the digital revolution, there remains a possibility that organizations have not fully transitioned towards harnessing the potential of smart data [10].

As organizations are encouraged to adapt to the emerging knowledge-driven economy, investigating the transformation from big data to smart data becomes essential. This investigation involves examining the differences and similarities between contemporary theoretical perspectives and practical experiences, providing insights into the evolving role of knowledge management in the digital era. By exploring this interplay, we can comprehensively understand how the digital revolution has influenced knowledge processes within manufacturing companies, thus establishing the theoretical framework to effectively address our research question.

In their empirical work, [6] conducted a questionnaire involving 162 Malaysian manufacturing firms, highlighting the significance of integrating knowledge management (KM) into business processes. They concluded that KM plays a crucial role in driving process- and product innovation, leading to innovative value creation. Their findings provide valuable insights and guidelines for effective KMPs.

The study defines four dimensions of KM (shown in Fig. 1), which serve as the foundation for our case study. These dimensions examine how organizational knowledge is: created and acquired; shared and transferred; stored and retrieved; and applied. Their findings emphasize the importance of interrelationships among KM dimensions for their overall effectiveness, increasing the firm's potential to leverage the inherent value of their data for effective innovations [6].

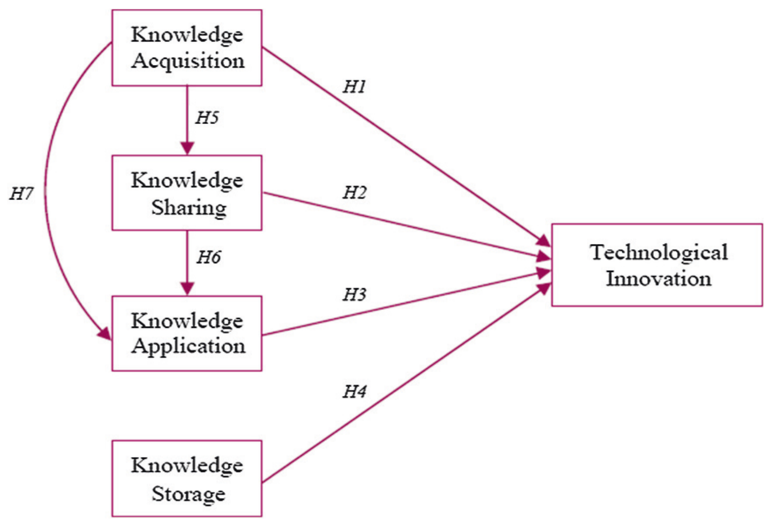

Fig. 1 Overview of the 4 knowledge management dimensions, their interrelationships and influence on technological innovation, according to [6]

[7] contributes to the prior industry experience by providing their academic perspective. They conduct a bibliometric analysis of 90 relevant articles and use cluster analysis to explore the intellectual structure of academic literature related to knowledge management in Industry 4.0. Their analysis results in the identification of six clusters of relevant keywords, which serve as the basis for their systematic literature review. Their research allows them to identify gaps in the existing academic literature and integrate current trends and emerging technologies, ideas, and concepts. This enables them to provide insights into the future directions that the knowledge management literature may take.

Therefore, by comparing the practical guidelines for applying knowledge management in industry with the state of knowledge management in academic research, it is possible to integrate emerging concepts and ideas related to KM. This integration allows for the identification of parameters that contribute to the development of a robust digitalization strategy applicable to SMEs in general, and specifically to Wiewald.

3.1 Knowledge Acquisition & -Creation

The academic sphere places significant emphasis on 'knowledge creation', with approximately 50% of the papers in the dataset addressing this topic [7]. The creation of knowledge involves the generation of new knowledge within the company or its acquisition from external sources. By adopting this broader definition of knowledge acquisition, its link to innovation is reinforced, as the creation of knowledge significantly enhances innovative performance [6]. Consequently, when internal and external knowledge sources are effectively combined, the acquisition of knowledge exerts a positive influence on innovation.

For this case study, 'knowledge acquisition' involves discovering, locating, creating, or capturing new information, including both intangible tacit knowledge and tangible explicit knowledge from diverse sources such as employees, data, and external entities [6, 7]. [7] primarily focusses on 'knowledge capture,' highlighting its positive relationship with 'knowledge sharing' and '-transfer', with guidelines promoting an effective knowledge-sharing culture that is impactful to decision-making and thereby relevant to the 'knowledge application' dimension [6].

'Knowledge creation' trends overlap with '-transferability' trends, such as smart and digital environments. This results in two sub-streams of academia; process modelling and condition monitoring, which prioritize cyber-physical- and semantic technologies like digital twins, semantic webs, and semantic web rule languages to enhance 'knowledge acquisition' capability [7].

3.2 Knowledge Sharing

'Knowledge sharing', also known as 'knowledge accessibility,' allows employees to share and access both tacit- and explicit knowledge within and outside the organization [7]. The presence of a culture of 'data-mindedness' acts as a moderating variable, enhancing the speed of information flow, thereby emphasizing the need for an integrated transparent knowledge-sharing behaviour [6].

This transparency creates a 'proximity effect', enabling increased access to information and faster knowledge dissemination during day-to-day operations [6]. Contem porary views highlight the paramount importance of 'knowledge sharing', evident in its prevalence across academic research and its influence on various trends [7].

Although the transfer of knowledge is often regarded as the most significant knowledge management dimension [6], its direct impact on strategic objectives, such as achieving higher AI-maturity, may be debatable.

3.3 Knowledge Storage

'Knowledge storage', or 'knowledge documentation,' is vital in the knowledge management process, involving capturing, refining, structuring, integrating, and storing information [6]. This concept resonates with academic topics such as big data, digital transformation, and cyber-physical systems that are heavily influenced by the digital revolution [7].

The concept of 'organizational memory' involves capturing tacit knowledge and enhancing its accessibility, thereby preventing the loss of valuable information [8]. This, in turn, fosters 'organizational agility' by enabling organizations to assimilate up-to- date knowledge and respond swiftly to new information. By combining these practices with the 'proximity effect' and fostering a culture of 'data-mindedness', knowledge storage practices accelerate the transmission and dissemination of knowledge, thereby driving innovative behaviour [6].

It is reasonable to assume that within such a dynamic environment, shaped by effective knowledge storage, that organizational knowledge can be leveraged to make well- informed decisions for long-term sustainable growth and an immediate competitive edge.

3.4 Knowledge Application

In the context of the global paradigm shift driven by the digital revolution, the importance of 'knowledge application', also known as 'knowledge responsiveness', becomes increasingly pronounced [6]. It represents a strategic asset for knowledge-driven organizations, facilitating a realignment of goals towards long-term strategic value creation. This realignment extends beyond the scope of business intelligence; which focuses on immediate data utilization, and enters the realm of data science. Through data science, organizations can generate profound insights that drive innovative advancements in business processes and products. Beyond supporting organizational goals, such strategic utilization of knowledge promotes collaboration and innovation, leading to the development of new products and processes. In fact, research has shown that effective knowledge application facilitates the translation of organizational expertise into innovative advancements, contributing to overall organizational innovation [6].

This transformative process towards advanced data-driven decision-making, includes the concept of 'responsiveness' that encompasses both 'agility' in reacting to changes in information and the 'quality of response'; which empowers organizations to take strategic actions and unlock new possibilities of higher quality, agility and effectiveness [6].

While the Malaysian industry questionnaire recognizes the strategic significance of 'knowledge application', ranking it as the second most important dimension in relation to innovation [6], the academic literature on KM reveals that it is the least investigated dimension, with only 16% of the papers in the dataset referencing knowledge application [7]. This discrepancy underscores the need to address this potential research gap and explore how the effective application of knowledge can bridge the gap between information advantages and actionable decisions.

Drawing insights from the investigative results of [5] and [7], we align our understanding with the current state of knowledge management practices (KMPs) in the field. These results emphasize the vital role of 'knowledge application' in achieving impactful outcomes in the digital era [3, 5]. It becomes evident that SMEs need to prioritize a digitalization strategy that effectively applies their knowledge resources to create a competitive advantage and foster innovation. By doing so, they can thrive in the rapidly evolving digital landscape. These findings underscore the significance of strategic knowledge management and its relevance in driving success for SMEs in the digital age.

4 Implication and Conclusive Recommendation for Wiewald

In partnership with Wiewald, we have been working on a compressor system 'configurator', that is supposed to assist their chief technician during the planning and design of new compressor systems. With early feasibility tests using historic data that was available to us at the time, we were able to design simple decision tree models that were able to determine most likely decision paths, when given basic requirements about desired compressor capacities. This initial work has sharpened our vision for the digital solution that suits Wiewald's requirements.

As the chief technician elicits the requirements from their customer, he enters the relevant conversation data into his electronic quotation form. The configurator uses this input to perform calculations in a discrete backend service. Once the basic datapoints are determined, the configurator should be able to provide suggestions for valid compressor systems.

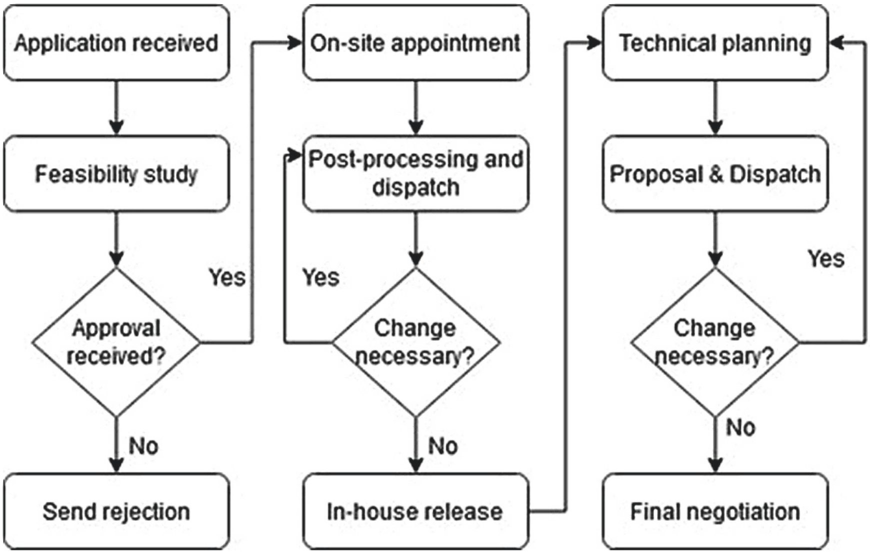

Fig. 2 Approximation of the current proposal creation process at Wiewald

The benefits of this configurator will be demonstrated by its rapid responsiveness and the quality of decisions it facilitates. With this prototype system, the customer should be able to gain an accurate understanding of the anticipated system price and its complexity before the session of the quotation process has concluded. Further adjustments to the requirements can be explored during the very same conversational session. This should reduce a two-week process with multiple tedious back and forth requirement-elicitation sessions (as shown in Fig. 2) into a single transparent session in which requirements and potential solutions can be communicated effectively and dynamically. Furthermore, the session can be concluded with the signing of a hand-over confirmation form. Requirements elicitation is central to their business process landscape; its successful completion determines and triggers all necessary business processes.

As of recent, we have received a rich and qualitative bounty of new data that we aim to utilize for the creation of a new prototype of superior sophistication. Our approach is inspired by the designs of the 'Actionable Cognitive Twin' (ACT) proposed by [9]. ACT establishes a solid foundation that enhances information transparency, decision support, and agility, while also ensuring that workers actively participate in the shaping of business processes, thereby driving innovative initiatives. This integrated approach aligns with the goal of effectively utilizing digitalized organizational domain knowledge and fostering a culture of data-mindedness and innovation:

- It captures both explicit and tacit expert knowledge within Wiewald, providing a means for systematic knowledge management and its explicit integration into their business process model; leveraging ontologies, knowledge graphs, and embeddings to externalize implicit information, thereby fostering a comprehensive understanding of the business processes. This integration of knowledge promotes effective ‘knowledge capture’ and ‘-preservation’, ensuring that organizational expertise is embedded within the company’s business operations.
- Furthermore, ACT serves as a holistic measure for innovation management, encouraging active involvement from workers, especially during the structuring and modeling of business data and processes, and facilitating worker-driven initiatives for innovation. This empowers employees to contribute their valuable insights and ideas, creating an environment that fosters collaboration and drives innovation within the organization. Beyond promoting worker empowerment, the implementation of ACT enables the adoption of AI-solutions to further enhance innovative organizational capabilities

To implement ACT, we would employ a data-driven approach, that leverages natural language processing (NLP) techniques such as entity recognition, relationship extraction, and semantic parsing. Through these techniques, we would extract relevant information from collected data and utilize neural network-based ontology learning methods to automatically learn ontologies from structured data and textual resources. These ontologies capture complex relationships and hierarchical structures, providing a foundation for knowledge representation. Furthermore, the ontological learning process would include the creation of an ontological embedding, which represents the distilled information from the learned ontology. This embedding serves as valuable information for an informed neural network (INN), guiding its understanding and reasoning capabilities.

The INN, equipped with the ontological embedding, is responsible for the capture and analysis of organizational data, leveraging the knowledge captured in the embedding to interpret and make sense of the incoming data. A knowledge graph serves as an instance of the learned ontology and is dynamically populated and updated by the INN. By populating the knowledge graph with relevant data from real-time sources (like the requirements elicitation form), the INN would create an interconnected representation of the acquired information.

The knowledge graph, shaped by the INN, provides a structured and comprehensive view of the information captured from the interactive data streams, representing an integrated view of the entire organization’s knowledge. Therefore, it should enable systematic knowledge management, facilitate decision-making, and foster innovation within the organization. Furthermore, continuous refinement and validation of the domain knowledge ontology, knowledge-graphs, and ontological embeddings, in collaboration with domain experts, will be needed to ensure their accuracy, relevance, and coverage.

By implementing the configurator as an instance of ACT; an integrated approach, combining ontological learning, ontological embedding, and INN-driven knowledge graph population, we aim to unlock new possibilities for effective knowledge representation, reasoning, and information retrieval for the task of compressor system configuration at Wiewald.

References

1. Shares of small and medium-sized enterprises in selected variables. (2020). https://www.destatis.de/EN/Themes/Economic-Sectors-Enterprises/Enterprises/Small-Sized-Enterprises-Medium-Sized-Enterprises/Tables/total-cik.html. Zugegriffen: 16. Jun. 2023.
2. SMEs Digital Strategies for the digital transformation. https://www.bmwk.de/Redaktion/EN/Publikationen/Mittelstand/smes-digital-strategies-for-digital-transformation.pdf?blob=publicationFile&v=5. Zugegriffen: 16. Jun. 2023.
3. The Art of AI Maturity | Accenture. https://www.accenture.com/us-en/insights/artificial-intelligence/ai-maturity-and-transformation. Zugegriffen: 16. Jun. 2023.
4. Xu, X. et al. (2021). Industry 4.0 and Industry 5.0—Inception, conception and perception. *61*, 530–535. https://doi.org/10.1016/j.jmsy.2021.10.006.
5. Leoni, L. et al. (2022). The mediating role of knowledge management processes in the effective use of artificial intelligence in manufacturing firms. *42*(13), 411–437. https://doi.org/10.1108/ijopm-05-2022-0282.
6. Lee, V. -H. et al. (2013). Knowledge management: A key determinant in advancing technological innovation? *17*(6), 848–872. https://doi.org/10.1108/jkm-08-2013-0315.
7. Manesh, M. H. K. et al. (2021). Knowledge management in the fourth industrial revolution: Mapping the literature and scoping future avenues. *68*(1), 289–300. https://doi.org/10.1109/tem.2019.2963489.
8. Bhandary, A., & Maslach, D. (2018). *Organizational Memory*. Palgrave Macmillan, UK. https://doi.org/10.1057/978-1-137-00772-8_210.
9. Rožanec, J. M. et al. (2021). Actionable cognitive twins for decision making in manufacturing. https://doi.org/10.48550/arxiv.2103.12854.
10. Das Gold des 21. Jahrhunderts – Digitale Daten aus Fahrzeugen. https://tinyurl.com/bde98pab. Zugegriffen: 16. Jun. 2023.

Manufacturing and factories of the future

Abstract. The papers of this section deal with developments and potentials for manufacturing industry that come with AI. The contributions highlight the production of the future and topics that shape the innovation process, such as Industry 4.0, the Internet of Things (IoT) and robot-assisted automation. Examples of the use of AI in production, process optimization and quality control are presented.

Identification of Machine Learning Algorithms to Share Tacit Experimental Knowledge in Manual Production

Christian Prange[1](✉), Amin Beikzadeh[1], Holger Dander[1] and Nicole Ottersböck[2]

[1] Lehrstuhl Fertigungstechnik, Universität Duisburg-Essen, Lotharstraße 1, 47057 Duisburg, Germany
{christian.prange,amin.beikzadeh, holger.dander}@uni-due.de

[2] ifaa – Institut für angewandte Arbeitswissenschaft e. V., Uerdinger Str. 56, 40474 Düsseldorf, Germany
n.ottersboeck@ifaa-mail.de

Abstract. This paper presents the research project "KI_eeper – Know-how to keep" and a first technical research. In this project, tacit experiential knowledge of employees is to be recorded and processed by using artificial intelligence to make it available to inexperienced workers due to a digital assistance system.

First, the difference between tacit and explicit knowledge is briefly explained. The existing case studies will then be examined in more detail and it will be explained why artificial intelligence is necessary for a general solution to identify and storage knowledge in this research project.

Subsequently, various machine learning models and algorithms will be considered, which could be used for a potential technical solution.

Zusammenfassung. Die Babybommer-Generation der 1960er Jahre wird in den kommenden Jahren verrentet. Dadurch verlieren Unternehmen langjährige erfahrene Beschäftigte und auch deren Wissen, wenn nicht effiziente Lösungen gefunden werden, dieses zu identifizieren, zu speichern und zu transferieren. Dieser Herausforderung widmet sich das Forschungsvorhaben „KI_eeper – Know-how to keep". Im Projekt wird erforscht, inwiefern künstliche Intelligenz Möglichkeiten eröffnet, um das implizite Erfahrungswissen von Beschäftigten automatisiert im Arbeitskontext zu erfassen, zu verarbeiten und zu transferieren. Aktuelle Ansätze des Wissenstransfers sind aufwendig und häufig auch mit hohen Kosten verbunden. Dabei hat jedoch eben das implizite Wissen große Relevanz für Unternehmen, welche für kleine und mittlere Unternehmen (KMU) noch stärker ausgeprägt ist.

Am Ende des Projektes soll ein digitales Assistenzsystem entstehen, welches das gesammelte und ausgewertete implizite Wissen von Erfahrungsträgern allen Beschäftigen zugänglich macht und diese somit bei der Ausführung ihrer Tätigkeiten bedarfsgerecht unterstützt. Dabei soll zunächst der Fokus auf Tätigkeiten in der Produktionsfertigung gelegt werden. Aufgrund der großen Vielfallt an Fertigungsverfahren und damit verbundenen unterschiedlichen

C. Zinke-Wehlmann und J. Friedrich (Hrsg.): *First Working Conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow*, AIDRST 2023, Informatik aktuell, S. 143–151, 2024. https://doi.org/10.1007/978-3-658-43705-3_11

Daten, soll eine allgemeingültige Lösung für Verarbeitung der Daten mittels der künstlichen Intelligenz gewählt werden. In dieser Veröffentlichung sollen ausgewählte Algorithmen betrachtet werden, welche für eine potenzielle technische Lösung verwendet werden können.

Keywords: Knowledge storage · Application and case study · Human-centered approaches of AI design & development · Artificial intelligence

1 Introduction: KI_eeper – Know-how to keep

The research project "KI_eeper – Know-how to keep", founded by the German Federal Ministry of Education and Research (BMBF), consists of a consortium of research institutions, developers and two companies from the metal and electrical industry in Germany. These provide case studies on which the technical solution aimed in the project is to be developed and tested. The aim of this project is to identify, safe and transfer the tacit experiential knowledge of long-term employees by using artificial intelligence.

Knowledge can generally be distinguished as tacit and explicit. Explicit knowledge is understood as knowledge that can be clearly communicated. It can be expressed in words and numbers, as well as passed on to others as, for example, data or processes [1]. Tacit knowledge, on the other hand, is knowledge that cannot be fully or adequately communicated in language by the outgoing person. This knowledge is often unknowingly embodied in the minds of employees and is manifested in intuitive behavior [2]. Activities do not always consist of explicit or tacit knowledge alone, but of the mixture of both. The more tacit knowledge is required, the more difficult it is to record and transfer knowledge to another person. Tasks with high proportion of tacit knowledge include activities with complex procedures and non-standardized processes [3], which can often also be found in manufacturing companies. According to North [4], a basic prerequisite for the competitiveness of a company is that employees can apply their experiential knowledge by training-on-the-job. If a long-term experienced employee leaves a company without passing the tacit knowledge, this can result in a loss of competitiveness. On the other hand, successful handling of existing knowledge in a company can even lead to an increase in competitiveness [5].

Knowledge management is concerned with how to transfer knowledge to other employees within an organization to prevent the disadvantages described by North. Approaches such as coaching, mentoring or training offer the opportunity to pass on tacit specialist knowledge [6]. Due to the demographic change, such concepts can often no longer be pursued, as many specialist positions are not filled and thus the direct transfer of the necessary knowledge is no longer possible. An increasing number of companies are having problems recruiting skilled workers, because the shortage of skilled workers is being acute in manufacturing companies [7]. In addition, the transfer of knowledge is hindered by language barriers of employees with migration background. Therefore, a solution should be found how valuable knowledge can be absorbed and stored as efficiently as possible. This valuable implicit knowledge should subsequently be made available to unskilled workers by means of assistance systems.

2 Use Cases

In the current context of demographic change, the aim of the research project is to record tacit knowledge and convert it into explicit knowledge which can then be presented to all employees. In particular, new inexperienced employees should be able to be trained more quickly in this way. This is to be elaborated based on two very different case studies. The use cases in the companies have been selected in such a way that the input and output for the intended assistance system differ greatly. In addition, the outputs of the assistance system should be selected in such a way that they are presented as simply as possible, so that people with language barriers or semi-skilled employees can also use the system and manage the working process professionally.

2.1 First Use Case: Straightening Flat Steel

For the first use case, an activity was chosen that requires a high degree of tacit knowledge on the part of an individual person. This involves the manual straightening of flat steel, which can be available in different dimensions and material properties. The raw material can be up to three meters long, hardened or non-hardened and weigh up to 25 kg [8]. For the straightening itself, an oscillating hydraulic joining press is used, in which only the ram adjustment is infinitely settable. The change in the ram adjustment results in a change in the force acting on the flat steel. According to their own statements, the experienced employees at this workstation cannot describe how to straighten the flat steel. They work intuitively, which is derived from their long-term experience at this station. They cannot convey why they use a certain force to straighten at certain points. They feel the rhythm of the press. They intuitively recognize which of the two edges can be straightened best. They adjust the pressure without having to look at the display. They can see with their eyes whether the flat bar still has a slight deflection. That is why it takes at least one years and more to learn this job.

The actual and target condition of the flat steel are used as the first input parameter. The target condition can be retrieved digitally from technical drawings. The actual condition of each flat steel is to be recorded with the use of a scanning sensor. For the training of the AI, the setting of the path limitation is also required. This can be retrieved digitally from the hydraulic press. Furthermore, the exact positioning of the flat steel in relation to the pressure point on the press is required so that the tacit knowledge can be processed by the AI. At present, a combination of distance sensors and the application and reading of a vernier is required for this. Thus, positioning is chosen as the third input variable for the flat steel. From these three input variables, the artificial intelligence is to calculate the correlation between the acting force and the pressure position on the bending of the flat steel. This correlation is to be made available to the inexperienced worker in the form of position data – again with the distance sensors and the vernier – as well as a setting suggestion for the path limitation.

Later, during the standard operation of the assistance system, only the actual and target states are to be used as input parameters. However, position data and path limitation settings are still to be defined as output parameters.

2.2 Second Use Case: Sub-Process of a Surface Technology System

The surface technology system of the other applications company was selected as the other use case. This is a plant process in which sheet metal parts are coated by means of powder coating and wet coating. In this process, the components are first clamped in a frame, which is then hung on a carriage of a rail system. This rail system transports the components to the individual stations of the system: cleaning, coating, curing, and hanging off with quality control. The complete system can be divided into several subsystems. The first station at the beginning of the process, where the components are hung up and the machine is controlled/operated, has the most influence on the efficiency and workload of the other subsystems. At this point, not only the frame is selected to match the components, but also the number of items, the positioning, the orientation of the components and the clamping elements used as all as the masking positions are determined. The masking serves to ensure that desired areas are not coated, e.g., threats or threaded holes.

The number of different coating products is higher than 2600 with a total of 250 various colors. Actual there is not much information/instruction on how the components are to be positioned on the frames, which auxiliary materials are to be used or which masking measures are to be taken. This information can be partially derived from existing technical drawings – including the masking measures. This results in the target state of the technical drawings as the first input variable for the training of the artificial intelligence. Furthermore, the AI must be fed with the experiential knowledge of the employees. For this purpose, an input is required where the employees enter the type of suspension, number of components, etc. In addition, a direct link between the AI and the technical drawings should be established at this point. In this case, a correlation to the recorded data as well as the target state should also be defined. As a third input variable, an image of the suspended components is to be recorded after the suspension process. This should be further associated with the other input variables. During the recording, not only the individual carriages are to be considered, but also their sequence. Since different sized components require various times during powder coating as well as other baking times, a certain sequence of diverse items must be adhered to maintain the desired cycle time. Currently, this planning of the sequence is only possible through the experience of longtime employees.

After the data training, the cognitive assistance system should provide all employees with suggestions regarding the hanging process when retrieving certain articles as an output. Furthermore, the sequence of articles to be processed should be determined and communicated by the artificial intelligence.

3 Methodological Approach

In order to find suitable AI models for the use-cases described, some criterion must first be selected so that a more detailed selection can be made.

At the time of this publication, the technical solutions only exist as concepts, which is why it is not yet possible to make a more precise statement regarding the sensor technology to be selected. However, IoT sensors will be used in both applications. Therefore, the selected model must be able to handle sensor data in general.

It was also determined that the model must be as transparent as possible or present data in such a way that an employee can cognitively grasp the learned properties of the model. Therefore, the internal processes of the model should have as little complexity as possible so that every employee can process the results of the AI. This results in complexity as a further criterion for selection. For the same reason, it must also be possible to process the interactions between the input variables when capturing the tacit knowledge. Considering these two prerequisites, it becomes clear that an explainable AI is being searched for, since the actual calculation processes running in the background are not available as a so-called "black box", but as a "white box". These processes are therefore available to the user and can be understood. This results in another positive aspect: from a socio-technical point of view, prejudices of employees could be reduced if it is recognizable what the actual model calculates or works with the data.

The scalability of the AI model also plays a major role. A generalist approach can result in both large and small amounts of data for different applications. Thus, the targeted AI must also be able to cope with large databases. In addition, accuracy is used as a criterion for the choice of model to obtain the highest possible resolution of the implicit knowledge.

To obtain a suitable selection of AI models, reference databases such as IEEE Xplore and ScienceDirect were accessed. These databases were first roughly searched for machine learning approaches which were rated using evaluation criteria. The generated data set was compared with the previously described criteria, e.g. that the model is explainable. Some of the resulting algorithms will be briefly presented in the following chapter.

4 Presentation of Different AI Models

Due to the variability of the two use cases, the biggest challenge in this project is to develop a universal AI-based solution, to capture the tacit knowledge of the experienced employees. Although in both cases there is digital data available for training that can be directly defined as input into the AI. The type of input of the experiential knowledge is very different. In the first case, the sensor system captures the knowledge completely passively. Whereas in the second use case, although certain data is passively recorded by taking photos and by an active input. The employees enter information into the technical system at the beginning of the technology introduction, e.g., using selection menus and images in one of the use cases. Over time, the system has collected so much data that employees have to actively enter less and less information. The system should then be able to make suggestions for improved task execution based on the collected data. In the following, some viable models that were identified due to research will be briefly presented.

4.1 Knowledge-Base Building Procedure

Huang et al. [10] addressed the involvement of the human expert knowledge in digital image processing using a machine learning method to build a knowledge-based model. They assume that each object can be represented by its attribute value class vector, such as [attribute a_1,…, attribute a_n, class – i]. Therefore, they chose a subset of data, S, as a representation of data, so that this subset of data includes all possible classes. Then, a decision algorithm is applied to the subset of data to partition S into subsets of S_1, S_2, …, S_n based on some production rules. These production rules express that the object can be assigned to a certain class if the rules are followed. For instance: if attribute 1 > 80 and attribute 2 < 10 then object will be belong to class 2.

4.2 GAMI-Net

The fundamental idea of the GAMI-Net described by Yang et al. [9] is to capture the main effects and pairwise interactions of the model inputs. From this point of view, the structure of the GAMI-Net consists of two disjoint networks "Main Effects Module" and "Pairwise Interactions Module", each performing its own tasks. The Main Effects Module tries to capture the nonlinear main effects of the model inputs using so-called subnetworks. Each of the subnetworks gets one input and propagates it through intermediate layers to calculate one output; while the pairwise interactions get two inputs to capture the partwise interactions of the model inputs. Hence, what distinguishes these two modules from each other is the number of input neurons in their input layers. At last, a linear combination of the modules outputs forms the final output. At last, a linear combination of the module outputs forms the final output. The following Fig. 1 depicts the structure of the GAMI-Net.

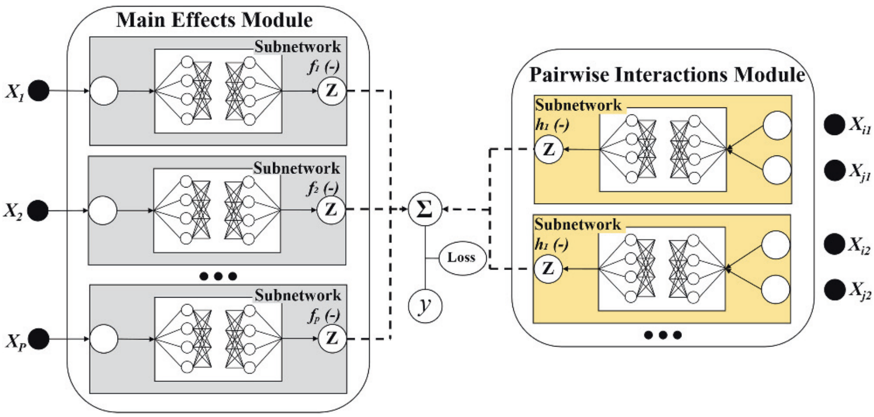

Fig. 1 GAMI-Net architecture based on [9]

4.3 Bayesian Rule List

The Bayesian Rule List by Letham et al. [11] is about the generation of a list of rules for classifying data. The model consists of two main steps:

1. Obtaining frequent patterns within the dataset using the FP-tree (frequent-pattern tree) algorithm.
2. Learning of a decision list by selection of the previously obtained patterns.

Thus, the first step is to learn frequently occurring patterns. The authors use the FP-tree algorithm as an example, but according to them it also works with all other algorithms that extract patterns in data sets. Accordingly in the further process a general representation of the finding patterns, which also happens with the FP-tree algorithm. Due to the more detailed explanation of FP-tree will be omitted. Patterns are here both the characteristics of a property within a data set as well as the combination of characteristics. The modeler determines when a data pattern occurs frequently.

4.4 Deep Learning for Case-Based Reasoning through Prototypes

Li et al. [12] attempted to present a neural network architecture (Deep Learning for Case-Based Reasoning through Prototypes – DLCBRP) through which without using a second model an explanation for each prediction could be possible. To do this, they proposed an encoder and a decoder. During training, the data set is processed by the encoder in such a way that the dimension of the input is reduced and useful properties for a prediction are retained. This reduced training data set is then used to learn prototype vectors. In this context, a prototype means that it is as close or identical to an observation of the training data set. The learned prototype vectors, whose validation might be visualized by an encoder network, reside in the prototype layer, which is then connected to the next fully connected layer by weights. Finally, by normalizing the weighted sum, a softmax layer output is given, which represents a probability distribution of all classes (Fig. 2).

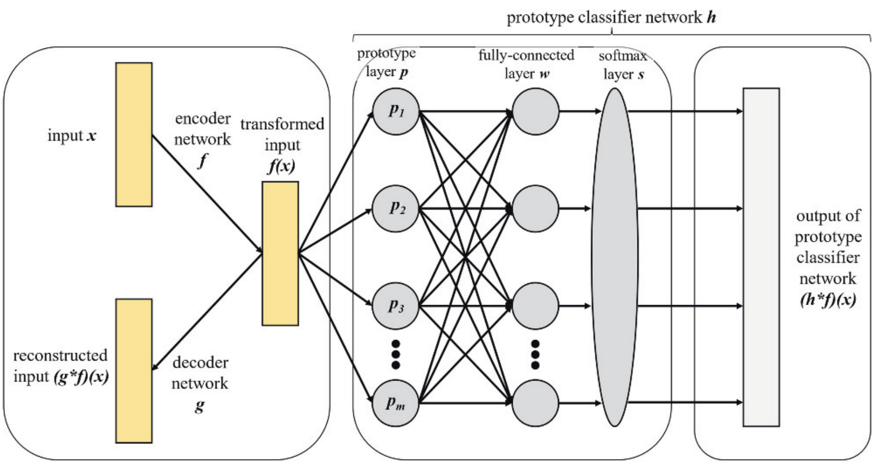

Fig. 2 DLCBRP-architecture based on [12]

5 Discussion

In the context of machine learning, not only the choice of an adequate approach should be addressed, but also the data that is fed into a model is most important. In the described cases, there is tacit knowledge that someone has acquired over years of working in a company. Collecting this tacit knowledge is one of the most challenging tasks, as any kind of sensitive knowledge must be taken into account in terms of the overall data quality. Hence, the model will learn from its training data more effectively if more correct training data can be provided. It should be considered that this kind of knowledge may not always can be expressed in form of data.

Another aspect that should be addressed is the explainability of the chosen models. After all, in the context of preparing or teaching someone who has not been familiar with a system, there is this expectation that after some hours or days they should be able to perform their work independently without the help of supervisors or others. Once a suitable model has been trained, it should be able to interpret which components or how much of each component is involved in a final product. This paper has therefore first tried to select some of the approaches that belong to the explainable models.

The AI models described before have different properties. However, none of the discussed algorithms provides a universally valid solution for the inclusion of tacit knowledge. With the paper, an intermediate goal has been achieved in which algorithms for the use case have been investigated. It is evident that further work is required to achieve the overall project goal.

Acknowledgements. The presented results are part of the research project „KI_eeper – Know how to keep" (FKZ: 02L20C500- 02L20C505). We would like to thank the German Federal Ministry of Education and Research (BMBF) for funding this project. Furthermore, we thank the Project Management Agency Karlsruhe (PTKA) for supporting the project. The authors are responsible for the content of this publication.

References

1. Nonaka, I., & Takeuchi, H. (1995). The knowledge-creating company. *How Japanese companies create the dynamics of innovation / Ikujiro Nonaka and Hirotaka Takeuchi*. Oxford University Press, Oxford.
2. Neuweg, G. H. (2001). *Könnenschaft und implizites Wissen: zur lehr-lerntheoretischen Bedeutung der Erkenntnis- und Wissenstheorie Michael Polanys*. Waxmann, Münster.
3. Implizites Wissen im Unternehmen. (2023). http://www.implizites-mitarbeiterwissen.de/implizites-mitarbeiterwissen/ (date of retrieval 03.04.2023).
4. North, K. (2016). *Wissensorientierte Unternehmensführung*. Wiesbaden: Wissensmanagement gestalten / Klaus North. Springer Gabler.
5. Diedering, R. (2018). *Wissensmanagement. Faktoren zum Erhalt des Wissens einer Organisation vor dem Hintergrund des demographischen Wandels*. GRIN Verlag.
6. INTERNATIONAL ATOMIC ENERGY AGENCY (2022). Mentoring and Coaching for Knowledge Management in Nuclear Organizations, IAEA-TECDOC-1999, IAEA, Vienna. https://www.iaea.org/publications/15089/mentoring-and-coaching-for-knowledge-management-in-nuclear-organizations.

7. Hardege, S. (2023). DIHK-Fachkräftereport 2022.
8. Ottersböck, N., Dander, H., Prange, C., Peters, S., & Ochterbeck, J. (2023). Flexibler Arbeitskräfteeinsatz durch KI-basierten Wissenstransfer. Praxisbericht aus dem Forschungsprojekt KI_eeper – Know how to keep. WERKWANDEL.
9. Yang, Z., Zhang, A., & Sudjianto, A. (2021). GAMI-Net: An explainable neural network based on generalized additive models with structured interactions. *Pattern Recognition, 120*, 108192. https://doi.org/10.1016/j.patcog.2021.108192
10. Huang X., Jensen J. (1997). A machine-learning approach to automated knowledge-base building for remote sensing image analysis with GIS data. *Photogrammetric Engineering & Remote Sensing, 63*(10), 1185–1194.
11. Letham, B., Rudin, C., McCormick, T. H., & Madigan, D. (2015). Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model. *The Annals of Applied Statistics, 9*. https://doi.org/10.1214/15-AOAS848.
12. Li, O., Liu, H., Chen, C., & Rudin, C. (2018). Deep learning for case-based reasoning through prototypes: a neural network that explains its predictions. *AAAI, 32*. https://doi.org/10.1609/aaai.v32i1.11771.

An Application of AI for Online Estimation of the Impact of Imperfections in Additive Manufactured Components

Denise Holfeld[1](✉), Franziska Theurich[1], André Rauschert[1], Gregor Neumann[2], Falk Hähnel[2] and Johannes Markmiller[2]

[1] Fraunhofer Institute for Transportation and Infrastructure Systems, Zeunerstr. 38, 01069 Dresden, Germany
{denise.holfeld,andre.rauschert}@ivi.fraunhofer.de
[2] Institute of Aerospace Engineering, Technische Universität Dresden, 01062 Dresden, Germany
{gregor.neumann,falk.haehnel, johannes.markmiller}@tu-dresden.de

Abstract. Artificial intelligence (AI) is popular for applications in image or natural-language processing, but AI can also be used to learn complex relations in production processes. For example, an AI can predict product quality based on process data during the production. In this paper, we present an application of AI to estimate structural properties of additive manufactured components in real-time. Occurring imperfections, such as air inclusions in the component, are considered and evaluated, since these have a significant influence on the quality of the component. This approach combines finite element (FE) simulation and machine learning: based on FE simulations, a neural network is trained to represent the relation between imperfections and the robustness of the component. To predict the impact of imperfection in real-time, monitoring systems are used to detect anomalies during the printing process, which are indications for imperfections in the additive manufactured component. Afterwards, the trained model is used to evaluate the impact of the detected anomalies to the component quality. This application of AI has a great potential to improve the additive manufacturing process itself and simplifying the approval of additively manufactured components.

Zusammenfassung. Die bekanntesten Einsatzgebiete von künstlicher Intelligenz (KI) sind zurzeit Bild- und Sprachverarbeitungen. Im Gegensatz dazu liegt der Fokus hier auf dem Lernen komplexer Zusammenhänge in Produktionsprozessen: untersucht wird ein KI-Anwendungsfall, in dem die Qualität des Endproduktes basierend auf Prozessdaten, die während der Produktion erhoben werden, vorhergesagt wird. Mit Hilfe Neuronaler Netze wird die Belastbarkeit von additiv gefertigten Bauteilen in Echtzeit bewertet. Dazu werden Imperfektionen, wie z. B. Lufteinschlüsse im Bauteil, berücksichtigt, da diese einen erheblichen Einfluss auf die Qualität des Bauteils haben. Das heißt, basierend auf Prozessdaten des additiven Fertigungsprozesses werden potenzielle Imperfektionen im Bauteil detektiert und in eine dreidimensionale

C. Zinke-Wehlmann und J. Friedrich (Hrsg.): *First Working Conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow*, AIDRST 2023, Informatik aktuell, S. 153–163, 2024. https://doi.org/10.1007/978-3-658-43705-3_12

Repräsentation des Bauteils eingefügt. Mit Hilfe des trainierten Neuronalen Netzes werden dann Festigkeitskennwerte für das Bauteil bestimmt, die Aufschluss über die Qualität geben.

Zum Trainieren eines Neuronalen Netzes ist ein großer Trainingsdatensatz notwendig. Diese Daten für diesen Anwendungsfall experimentell zu generieren würde bedeuten, dass sehr viele Bauteile mit Imperfektionen gefertigt und anschließend geprüft werden müssten. Um diesen zeitaufwendigen und unwirtschaftlichen Prozess zu umgehen, werden Finite-Elemente-Simulationen zur Erstellung der Datenbasis genutzt. Das heißt, es werden Finite-Elemente-Modelle des Bauteils erzeugt, in die Imperfektionen künstlich eingebracht werden. Mit numerischen Simulationen können anschließend die Festigkeitskennwerte des Bauteils ermittelt werden. Da diese Simulationen zeitaufwändig sind, ist eine Echtzeit-Anwendung nicht möglich. Stattdessen trainieren wir mit den Ergebnissen der Simulationen ein Neuronales Netz, dass die Simulationsergebnisse vorhersagen soll. Ein solches Neuronales Netz kann dann während des Produktionsprozesses genutzt werden, um in Echtzeit die Auswirkungen der detektierten Imperfektionen auf die Qualität zu beurteilen. Dadurch können mangelhafte Bauteile aussortiert und eine gleichbleibende Qualität gewährleistet werden. Es ist sogar möglich, den Druck mangelhafter Bauteile abzubrechen, was Zeit und Ressourcen spart.

Keywords: Machine Learning · Neural Networks · Finite Element Method · Additive Manufacturing · Quality Assurance

1 Introduction

Machine learning and artificial intelligence (AI) applications have evolved significantly in the past few years. One field of application is finding relations in large datasets and learning the underlying unknown complex mathematical functions. In this paper, we present an AI-based approach that enables quality assurance in an additive manufacturing (AM) process. AM, or 3D printing, is a category of production processes where components are created layer wise. It enables the production of highly complex components of any geometry and even the design of internal structures such as cooling channels or injection nozzles. This design freedom allows to produce lightweight and cost-effective components, which makes this technology very attractive for the aerospace industry. Since AM is a relatively new technology, there is much less experience compared to traditional manufacturing processes, e.g., in terms of the impact of disturbances in the production process on the component quality. Consequently, in order to be allowed to install an AM component, e.g., in an aircraft, an approval process must be passed consisting of a lot of destructive and non-destructive tests. More precisely, destructive tests are required to verify general material properties and the strength of a component. Further, non-destructive testing is required to examine each component individually and, if necessary, to initiate post-processing steps for quality assurance. For example, a computer tomography (CT) scan is used to detect imperfections. Unfortunately, these needed tests and post-processing steps are very time-consuming and expensive. To reduce this disadvantage of the AM processes, we will show in this paper how production data obtained during the printing process can be used to assure component quality which will decrease the demand of tests and make the approval process more efficient.

2 Technology Background and State of the Art

In this paper, we specifically refer to the AM process laser powder bed fusion (LPBF), where components are produced layer wise by melting a fine-grained metal powder with a high-power laser beam: A three-dimensional component design is divided into several layers of a specific thickness in z-direction. For each layer, a uniform powder bed is created by the recoating process. The laser beam is directed over this powder bed with special scanning mirrors to scan the component areas in the current layer, where it melts the powder into a solid metal. The quality of AM components depends mainly on the metal powder used and on the printing parameters such as laser power, hedge distance, laser speed or layer thickness. However, local defects can occur even with well-chosen parameter settings, cf. [1]. The resulting imperfections have been shown to cause a negative change in the mechanical structural properties of the final component compared to the properties of an optimal component. For example, porosity contained in the AM component, caused by keyholing or lack of fusion, affects dynamic and static properties and can cause anisotropic material behavior, see [2, 3]. Furthermore, the surface roughness of AM components particularly affects the fatigue properties, cf. [4]. Consequently, comprehensive quality assurance during the production process is essential for the outcome of the AM process. To achieve this, monitoring systems are used to collect information about the manufacturing process from which production-related component defects can be derived. The challenge is to determine the influence of defects on the properties of the manufactured components. For this purpose, we present an approach to evaluate the influence of defects to the component quality in real-time using AI.

A crucial point in AM using LPBF is the stable melting of the metal powder. As shown, e.g., in [5], the temperature in the melt pool has a significant influence on the microstructure of the AM components. For this reason, various monitoring systems have been developed to observe the melt pool and several commercial suppliers of 3D printers offer monitoring systems for the melt pool or the thermal exposure, such as EOS, SLM Solutions, TruPrint, Renishaw and Sigma Additive Solutions.

We consider in this paper the optical tomography described in [6]. With this monitoring approach, the radiation during the LPBF process is recorded by an off-axial camera system, i.e., the camera does not follow the optical path of the fusion laser. The light omitted during the melting process consist of three components: reflected laser radiation, plasma emission and thermal radiation. A suitable band-pass filter is used to record only the thermal radiation. The layer-wise recorded data is visualized color-coded to show hot-spots and cold-spots. An analysis of AM test specimens has shown a correlation between the extreme temperature of spots and material failures.

Process data generation, i.e., the monitoring of the printing process, has already found its way into the industry, whereas the automatic detection of imperfections and defects is still the subject of current research. To provide an overview of this ongoing research activities, we state in the following a few examples where AI methodologies have been applied to identify anomalies in the AM process.

In [7], an application of supervised machine learning to detect defects is demonstrated. The approach is based on images taken for each layer with different lighting settings. Features for each location are extracted from these images and classified into anomalous and nominal. In this way, defects in the AM components can be detected and localized. A support vector machine is used for the classification. To generate the required reference data for supervised learning, anomalies of an AM component are detected from CT scans and the associated monitoring images are labelled.

In [8], three machine learning approaches are applied to detect hot spots based on high-speed videos of the printing process. Aim is, to classify between normal and defect-related brightness evolution. For this purpose, features were extracted from the videos, e.g., the mean gradient of the brightness in successive images, the maximum mean brightness drop between two consecutive frames, and the shape and size of the bright region. Then, these features are used to classify the video snippets by a support vector machine or by a neural network. Additionally, k-means is applied to cluster the development of the mean brightness. All three methods have shown an accuracy of more than 90 %.

However, the research results just listed are always only a detection of defects or anomalies. An evaluation of their impact on the component's quality is missing. Our vision is to evaluate the component's quality in real time based on the detected defects applying machine learning approaches. For this purpose, we need for many components information about their defects and their quality in terms of structural properties. To determine the structural properties of a component, it must be proven by destroying tests. For example, in tensile testing the component is slowly extended until it breaks. With it, e.g., the maximal force that the components can withstand is determined. However, this experimental approach requires to produce and destroy a large set of components to generate a representative database for training AI methods. This is very expensive and time-consuming and practical not feasible. To reduce this data generation effort, it was investigated, in a more general application context, whether numerical simulations can be used to create a representative training database. In particular, in [9], it is shown how deep neural networks can be used to estimate biaxial stress-strain curves of sheet metals. Based on synthetic generated crystallographic texture data, biaxial stress-strain curves were obtained from crystal plasticity-based numerical biaxial tensile tests obtained with finite elements (FE) simulations. With these numerical test data, deep neural networks were trained. It is shown that their predictions are close to the numerical computed test results. However, it is important to create a well-grounded data base to applicate the approach to real sheet metals.

In the following, we will transfer this idea to the AM production process. We present an approach that uses numerical simulations to generate a representative training database for AI methods. In detail, we developed an enhanced FE simulation that considers the material properties of AM components and allows to evaluate the influence of imperfections on the component quality. Note that the FE simulations itself could not be applied in a real-time quality assurance process because of the large computational time needed for one simulation.

3 Approach

The monitoring systems currently in use help to manually detect anomalies in the melting process which indicate the formation of imperfections in AM components. As mentioned above, several commercial 3D printers have integrated monitoring systems to record the melt pool, or the temperatures of the heat affected zone. These systems are an important enabler to detect imperfections. To gain greater benefit from monitoring, the detected imperfections must be analyzed in terms of their potential impact on component quality. The approach presented here provides such an evaluation.

Currently, the application of numerical simulation methods, like the FE method, is being investigated to determine structural properties of defect-afflicted components. In general, numerical simulation methods are based on a discretization of the component to numerically solve the partial differential equations that describe the behavior of a component under structural load. However, to obtain reliable results, a fine discretization is required. For this reason, these numerical methods are computationally expensive and cannot be used for a real-time service.

Another investigated method for monitoring and quality assurance of AM is machine learning. One problem of the application of AI methodologies is that usually a large number of data sets is required to train an AI. For AM applications, this means that many parts must be printed, especially parts with defects, to obtain a representative data basis. In other words, to train an AI that learns the relationship between defect position and structural properties, information about the component condition as a result of the manufacturing process (e.g., presence of imperfections) as well as the mechanical properties of the components are required. To obtain the mechanical properties experimentally, destructive tests must be carried out. Therefore, specimens or components have to be manufactured and manually tested one by one. This effort, i.e., manufacturing and testing a large number of components to obtain a sufficient data set to apply an AI method, is not economically feasible.

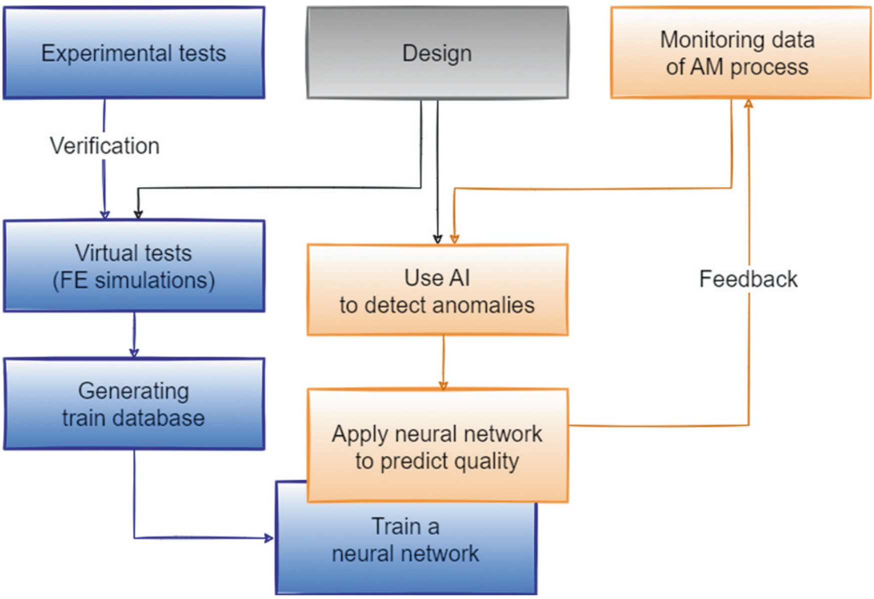

Fig. 1 Concept for AI based quality assurance in AM

In this paper, we combine these two approaches. The use of component-related and validated numerical simulation methods offers an alternative to the expensive production and testing step. In our approach, the training data is generated by means of FE simulations which compute the mechanical properties of components with imperfections. The data obtained will be used to train a neural network that can predict mechanical properties in real time based on the detected imperfection.

This approach is illustrated in Fig. 1 and consists of two steps: training a neural network to estimate the quality of AM components (blue colored) and an online application of the trained model during the printing process to predict component strength (orange colored).

3.1 Training of an AI Model

As it can be seen in Fig. 1, developing a neural network for quality prediction of AM components consists of three sub steps: virtual testing, generating training data and training itself. Furthermore, experimental tests are used to verify the results of the virtual tests.

Numerical simulation methods, i.e. FE simulations, are applied for virtual testing of a large set of components with various imperfections scenarios. In detail, the design of the component is discretized by defining finite elements, e.g., hexahedral volume elements. Based on this discretization, the deformation and stress under load are calculated by the numerical solution of partial differential equations. For this purpose, the material characteristics must be known, which can be obtained from experimental tests. One possibility to compute the effect of pores in the simulation is to define a FE model with pores by generating very small finite elements around the pore. However, since the pores in AM components are very small, with diameters around 10 µm [12], a very fine discretization is needed to represent pores which would lead to an extremely high computational effort. To reduce the overall computational effort, a multiscale method was used [10, 11]. For this purpose, representative volume elements are defined for a small piece of material with a pore. The corresponding material characteristics are computed by preliminary simulations based on very small finite elements. These representative volume elements are integrated in the FE model of the construction with larger finite elements and represent the imperfections. Subsequently, numerical simulations are used to predict structural properties and failure behavior of the virtually generated AM components with imperfections. However, due to the inhomogeneous material properties the computational effort is still high. With our developed FE model generator, the introduction of the imperfections as well as the generation of the simulation models can be easily realized. This offers the possibility to generate component models with arbitrary and specifically introduced imperfection states and to simulate their effects on the structural properties of the component.

To train a neural network, the FE models with imperfections must be provided as training data in combination with the virtual structural test results obtained from the simulations. To transform the FE models into input data for a supervised machine learning algorithm, the hexahedral volume elements have to be linked to pixels of a 3D image of the component design. This step is needed, because the elements of the

FE model have different sizes and shapes. Furthermore, they are described by their corners, which is not intuitive for finding relations between geometrical located defects and structural properties. In the generated 3D image, the pixels are associated with material properties to represent the FE model. For elements affected by defects, the material properties of the corresponding representative volume element are used for the corresponding pixels. A supervised learning approach is used, where the 3D images extracted from the FE models with imperfections are the inputs and the calculated structural properties of the FE simulation are the outputs. With these input and output data, a convolutional neural network was trained to represent the influence of imperfections to the structural properties.

In further research, experimental testing will be used to validate the virtual testing methodology. Note, that this validation requires much less experimental tests than generating training data for an AI only with those tests.

3.2 Online Application of the AI to Predict the Impact of Imperfections in Additive Manufactured Components

As indicated by the orange boxes in Fig. 1, after training a neural network, it can be used in the online quality monitoring. This step is also divided into three sub steps: Recording monitoring data, detection of anomalies by analyzing monitoring data, and estimation of the impact of the detected imperfections by applying the trained neural network to predict component's quality in terms of structural properties.

During the printing process, monitoring data is recorded, e.g., layer-wise monitoring images, which provide information about the temperatures in the heat affected zone, as shown in [6]. The recorded monitoring images have to be analyzed to detect anomalies in the printing process which are indications for imperfections like pores. Currently, different imaging methods are implemented in the 3D printers, but the monitoring data is often not analyzed automatically in real-time. Therefore, an image analysis tool needs to be developed to detect anomalies in the provided monitoring images. This can be either an AI, e.g., another neural network for image classification, or an image processing algorithm. After the automated detection, the anomalies must be localized in the printed components.

Based on the construction, a 3D image of the component is created and the detected imperfections are localized on the image. Based on this input data, the trained neural network predicts the structural properties. If the prediction does not meet the requirements, the AM component can be marked as insufficient or even the printing process can be aborted, which saves time and energy.

4 Feasibility Study

An initial feasibility study with tensile specimens indicated that the developed concept can be successfully implemented. First, training data for the machine learning step was generated using an FE model of a flat tensile specimen, as shown in Fig. 2. The hexahedral elements of the FE model are illustrated by the segmentation of the specimen. As it can be seen in Fig. 2, the elements in the examination area, where the specimen is thinner and the failure will occur, are significantly smaller than in the shoulders, where the specimen is gripped by the testing machine. Each element of the FE model is afflicted with material properties, which are amongst others Elastic Modulus, breaking point and Poisson's ratio. The elements with pores have different material properties because a pore reduces the strength. The example of Fig. 2 shows a FE model with two sizes of pores coloured in red and green.

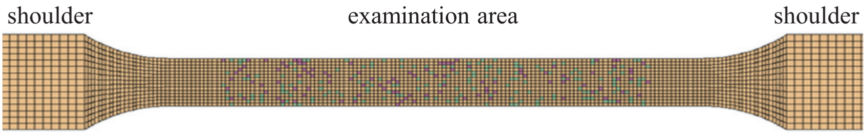

Fig. 2 FE model of a tensile specimen with imperfections (red and green)

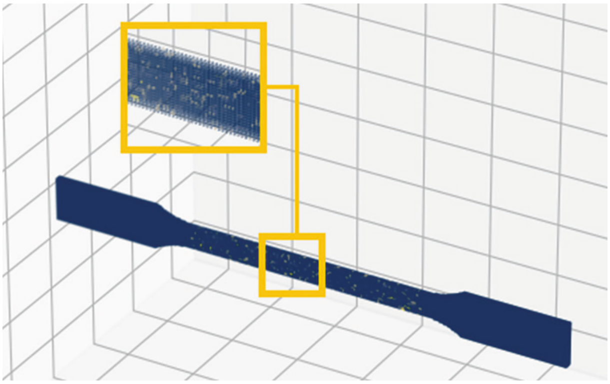

Fig. 3 3D-image representation of FE model

In the FE simulation, a virtual tensile test is done whereby the material properties of the elements affect the load distribution during the test and with it the strength of the whole specimen. One result of the virtual test is the Elastic Modulus for the specimen which shows the resistance to being elastically deformed in case of stress. For our initial feasibility study, we have used only this result of the FE simulation.

To train a neural network for the relation between pores in a specimen and Elastic Modulus of this specimen, a representative database is required. To generate the training data, several scenarios with different kinds of pores and distributions of pores are assumed, i.e., a small set of large pores, several small pores, or pores only in the shoulders of the test specimen. For these different scenarios, FE models with randomly distributed pores are generated and virtually tested. With it, the influence of the size and the distribution of the pores on the strength of the specimen is represented in the training data.

As mentioned above, for the neural network, the FE model is transformed into a 3D image of the specimen, as shown in Fig. 3, to serve as input data. However, each pixel represents the physical characterization of the corresponding element of the FE model. In Fig. 3, the dark blue coloured pixels show the material properties of perfect printed material, whereas the yellow-coloured pixels represent areas with imperfections and consequently with weaker material properties. In this way, the size difference between the finite elements can be taken into account and the localisation of pores as well as the closeness to other pores is represented.

These 3D images and the simulation results, i.e., the Elastic Modulus, which measures the stiffness of a material, are the training data. Thereby, the images are the input, and the Elastic Modulus is the output, because the neural network should learn, how pores in the AM component affect this physical property of the component. Due to the input is given as 3D-image, we used 3D convolutional layers for our neural network. After testing several training architectures and hyperparameters, a suitable neural network was obtained. The final architecture of the neural network consists of several successive units of one 3D-convolutional layer, one max-pooling layer and one batch normalization layer; a layer to flatten the obtained feature tensor; and a dense layer with 5 nodes as last hidden layer. With it, about 4000 trainable parameters must be determined. Furthermore, our investigations showed that normalization of the input data to [0,1] was more suitable than a standardization. For the training, we had about 350 simulation results. To enlarge the dataset, the specimens are rotated and mirrored, which will not affect the simulation results but results in more training data.

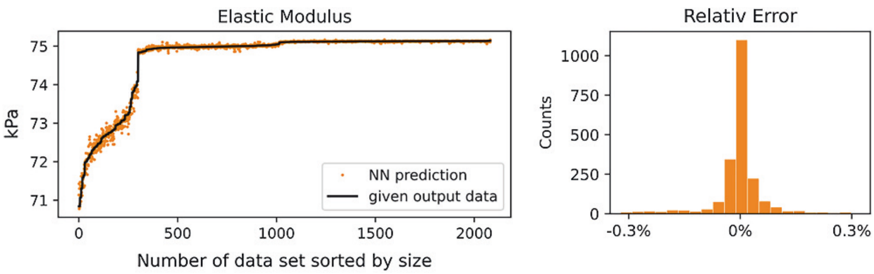

Fig. 4 Visualization of neural network's performance

Figure 4 visualizes the performance of the trained neural network. On the left, the prediction and the given output data are compared. Therefore, the data set is sorted by the output value, i. e. the Elastic Modulus computed by the FE simulation (black line); the orange dots show the corresponding predictions. As it can be seen, the predictions are close to the given output values and the general influence of the imperfections seems to be represented. On the right of Fig. 4, the distribution of the relative error, i.e., the difference between the prediction and the given output value in relation to the given output value, is shown. The relative error is generally small. In detail, the median of the relative error is −0.001 % which is close to zero. And for 95 % of the data points, the relative error was less than 0.21 %. The maximal relative error was 0.83 %.

5 Conclusions

In this paper, we presented an approach to apply AI in quality assurance of an AM process. Thereby, complex FE simulations were used to evaluate the effect of imperfections (i.e., pores) on structural properties of an AM component. To make this important information available in real time, the two disciplines FE simulation and machine learning have been combined. More specifically, a neural network is trained to represent the relations between imperfections that occur during production and the structural properties of the component. This neural network is used as predictor to estimate online the impact of the detected anomalies on the robustness of the component.

In the future, the data base will be expanded to include more complex component structures. Here, too, validated simulation tests will be used to determine the influence of pore frequencies and positions in the component on its quality. With the increasing data base, it is expected that not only correlations between pores and quality in concrete components can be predicted. With more complex neural network structures, more general relationships between, e.g., distances of pores to outer walls, pore sizes and wall thicknesses should also be learned.

We see great potential in this approach for improving the use of AM components. The experience gap compared to conventional manufacturing processes is reduced by data-driven approaches. Furthermore, since the approach can be validated by real material and component testing, it is possible to reduce the testing effort for approval in the future. In this case, significant material and cost savings can be expected.

The AM process itself can also be improved using this approach. For example, parameters can be adjusted or an AM process can be stopped if the AI model predicts that the AM component would not pass the approval process.

Acknowledgements. We acknowledge the financial support from the German Federal Ministry for Economic Affairs and Climate Action for the projects AMCOCS (Additive Manufactured Component Certification Services) and CertiFlight (End-to-end digital quality assurance for innovative approval processes based additive manufacturing technologies).

References

1. Gu, D., Meiners, W., Wissenbach, K., & Poprawe, R. (2012). Laser additive manufacturing of metallic components: Materials, processes and mechanisms. *International Materials Reviews, 57*(3), 133–164.
2. Brandão, A. D., Gumpinger, J., Gschweitl, M., Seyfert, C., Hofbauer, P., & Ghidini, T. (2017). Fatigue properties of additively manufactured AlSi10Mg – surface treatment effect. *Procedia Structural Integrity, 7*, 58–66.
3. Wang, P., Lei, H., Zhu, X., Chen, H., & Fang, D. (2019). Influence of manufacturing geometric defects on the mechanical properties of AlSi10Mg alloy fabricated by selective laser melting. *Journal of Alloys and Compounds Bd. 789*, 852–859.

4. Chastand, V., Tezenas, A., Cadoret, Y., Quaegebeur, P., Maia, W., & Charkaluk, E. (2016). Fatigue characterization of Titanium Ti-6Al-4V samples produced by Additive Manufacturing. *Procedia Structural Integrity, 2*, 3168–3176.
5. Mertens, A., Delahaye, J., Dedry, O., Vertruyen, B., Tchuindjang, J. T., & Habraken, A. M. (2020). Microstructure and Properties of SLM AlSi10Mg: Understanding the Influence of the Local Thermal History. *In Procedia Manufacturing, 47*, 1089–1095.
6. Zenzinger, G., Bamberg, J., Ladewig, A., Hess, T., Henkel, B., & Satzger, W. (2015). Process monitoring of additive manufacturing by using optical tomography. *AIP Conference Proceedings, 1650*(1), 164–170. American Institute of Physics.
7. Gobert, C., Reutzel, E. W., Petrich, J., Nassar, A. R., & Phoha, S. (2018). Application of supervised machine learning for defect detection during metallic powder bed fusion additive manufacturing using high resolution imaging. *Additive Manufacturing, 21*, 517–528.
8. Bugatti, M., & Colosimo, B. M. (2022). Towards real-time in-situ monitoring of hot-spot defects in L-PBF: A new classification-based method for fast video-imaging data analysis. *Journal of Intelligent Manufacturing, 33*, 293–309.
9. Yamanaka, A., Kamijyo, R., Koenuma, K., Watanabe, I., & Kuwabara, T. (2020). Deep neural network approach to estimate biaxial stress-strain curves of sheet metals. *Materials & Design, 195*, 108970.
10. Heinze, S., Joulaian, M., & Düster, A. (2015). Numerical homogenization of hybrid metal foams using the finite cell method. *Computers & Mathematics with Applications, 70*(7), 1501–1517.
11. Bargmann, S., Klusemann, B., Markmann, J., Schnabel, J. E., Schneider, K., Soyarslan, C., & Wilmers, J. (2018). Generation of 3D representative volume elements for heterogeneous materials: A review. *Progress in Materials Science, 96*, 322–384.
12. Kok, Y., Tan, X. P., Wang, P., Nai, M. L. S., Loh, N. H., Liu, E., & Tor, S. B. (2018). Anisotropy and heterogeneity of microstructure and mechanical properties in metal additive manufacturing: A critical review. *Materials & Design, 139*, 565–586.