Lausberg, Isabel (Ed.); Vogelsang, Michael (Ed.)

Book

AI in Business and Economics

Reference: (2024). AI in Business and Economics. Berlin : Boston : De Gruyter.
https://www.degruyter.com/isbn/9783110790320.

AI in Business and Economics

AI in Business and Economics

Edited by
Isabel Lausberg and Michael Vogelsang

DE GRUYTER

ISBN 978-3-11-079005-4
e-ISBN (PDF) 978-3-11-079032-0
e-ISBN (EPUB) 978-3-11-079043-6
DOI https://doi.org/10.1515/9783110790320

Library of Congress Control Number: 2024937910

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the internet at http://dnb.dnb.de.

Cover image: antoniokhr/iStock/Getty Images Plus
Typesetting: Integra Software Services Pvt. Ltd.
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Preface

The economic significance of artificial intelligence (AI) is accelerating. Therefore, we are pleased to present this collection of academic papers stemming from contributions and stimulating discussions at the Economic Perspective of Artificial Intelligence (EPEAI) conference held at the Ruhr West University of Applied Sciences (HRW) in Mülheim an der Ruhr, Germany, in March 2023.

The conference served as a platform for experts from different disciplines to explore the multifaceted intersection of economics and artificial intelligence. The book assesses the state of AI development, deployment, and the economic consequences. The authors use different methodologies on selected topics, analysing individuals as well as businesses and national economies. Overall, three conclusions can be drawn:

1. The implementation of a broad range of AI technologies in companies has just begun.

Markus Harlacher and colleagues explore how to implement AI in projects and how to select appropriate AI methods. Markus Feld, Wolfgang Arens-Fischer, and Marcel Schumacher argue that SMEs should follow a low-threshold approach when developing their own AI solutions. AI as a Service provides an easy access to AI technologies and is analysed by Nicolai Krüger, Agnis Stibe, and Jacqueline Krüger. Some contributions focus on specific business functions such as marketing: Khuliso Mapila and Tankiso Moloi summarise the challenges and opportunities of AI adoption mentioned in literature. Robert Menger, Karla Ohler-Martins, Amanda Lemette, and Robert J. Martin test how AI can classify the vast amount of documentation files for a power plant. Three papers shed light on accounting, reporting and business narratives: Thomas Rautenstrauch, Janis Hummel, Oliver Isoz, and Simon Moser highlight the consequences of automation for accounting. Isabel Lausberg, Anne Stockem Novo and Arne Eimuth propose a maturity model outlining a path to an AI-enabled management reporting and Irina Simon points out the relationship between AI and business data narratives. Finally, Marcus Bravidor discusses the challenges of integrating algorithmic audits with traditional audits and suggests frameworks to address these challenges.

2. The consequences of AI for employees and society are ambiguous.

Ed Dandalt explores the role of physicians and argues that they will only support AI when it does not threaten their job security. Timm Eichenberg, Nils Pudill, Britta Rüschoff, Anne Stockem Novo, and Michael Vogelsang shift the focus to job vacancies, identifying ideal-type personas of employees searched for in AI related job postings.

Two papers outline that a positive perception facilitates the transformation induced by new technologies: Zunera Rana, Jessica Roemer, Thomas Pitz, and Joern Sickmann explore newspaper articles about AI in Germany and find out that more

https://doi.org/10.1515/9783110790320-202

positive than critical articles were published. Simone Roth and Medina Klicic discuss generational differences in attitudes and suggest the use of a framing approach to increase intention to use social robots.

3. The adoption of AI encompasses the application of new methods, changing perceptions and the need to modify rules and regulations.
New methods to analyse data enable advanced predictions. Three teams compare deep learning models with respect to the demand of water, oil and stock markets: Katharina I. Köstner, Bàrbara Llacay, and David Alaminos use a Deep Autoregressive Recurrent Network (DeepAR) for predicting volatility in the oil market. Jan Vogt, Alexander Bönner, Michael Römmich, Malte Weiß, and Merih Türkoglu forecast prices of energy stocks using several Machine Learning methods including a LSTM and sentiment analysis. Gregor Johnen, Jens Kley-Holsteg, André Niemann, and Florian Ziel present a non-linear deep learning model which offers comprehensive probabilistic forecasts to measure the continuing effects of climate change on water usage.

Finally, AI adoption is also a matter of economic policy. Richard von Maydell and Christoph Menzel advocate for a modernisation of competition policy to prevent emerging competition deficiencies.

As can be seen from this brief overview, the book offers a variety of different contributions and perspectives. At the same time, it is clear that the research on the economic and business perspective of AI is only in its beginnings and there is a considerable amount of research to be done in the future.

We thank the authors, the conference participants, reviewers, and the De Gruyter team for making this anthology possible. Additionally, we sincerely thank the Scientific Committee members, Justus Haucap, Amanda Lemette, Tobias Meisen, Tankiso Moloi, Anne Stockem Novo, and Christian Weiß, for their hard work and commitment in selecting the conference papers. Special thanks go to Joshua Zander, whose organisational genius and exceptional dedication contributed in so many ways to make the conference a success. We would also like to thank Susanne Staude, President of the Ruhr West University, for her support and the many colleagues at HRW who joined us, especially Kathi Mulder and Björn Bovenkerk, who stepped in to cover for the hybrid event.

The responsibility for the accuracy and adherence to academic standards of the contributions contained within this volume rests solely with the authors of each individual work. Any views, findings, conclusions, or recommendations expressed in these papers are those of the authors and do not necessarily reflect those of the conference organisers, the editors or their institutions.

We are confident that this volume will serve as a valuable resource for scholars, practitioners, and students navigating the complexities of the economic implications of AI. The book provides approaches and methods to explore artificial intelligence from an economic perspective, a topic that will be ever more important in the future.

Isabel Lausberg and Michael Vogelsang
University of Applied Sciences Ruhr West
Mülheim an der Ruhr, Germany; March 2024

Contents

Part 4: **Organisation and Workflow**

Part 5: **HR and Employment**

Part 6: **Artificial Intelligence and Humans**

Part 1: **Competition and Regulation**

Richard von Maydell, Christoph Menzel

Chapter 1
The Rise of Artificial Intelligence: Towards a Modernisation of Competition Policy

Abstract: We aim at investigating to what extent the increasing importance of software, data and artificial intelligence (AI) may pose a potential threat to competition and requires further supervision and regulation of anti-trust and competition policy. Today, higher market concentration and higher markups can be observed in industries with high investments in intangible assets. Yet, no clear conclusions can be drawn by now about the specific effects of the increasing importance of software, data and AI on competition due to a lack of an appropriate database. Nonetheless, we suggest that political funding is needed to support the entrepreneurial integration of AI, as well as a modernisation of competition policy and anti-trust legislation in order to prevent emerging competition deficiencies in increasingly digitally-intensive economies at an early stage.

Keywords: artificial intelligence, intangible assets, markups, competition policy

1.1 Introduction

Already today, an increase in market concentration is noticeable in the markets for software, social media or communication networks, especially due to the rapid rise of Tech Giants, such as Apple, Alphabet, Meta, Microsoft or Amazon. Novel political and economic challenges need to be faced in increasingly digitally-intensive markets, which have already led to new competitive regulations such as the Digital Market Act (DMA), American Innovation and Choice Online Act, or the Digitalization Act as an amendment of the German Competition Act (GWB). The rise of the Tech Giants is largely based on progress in information and communications technologies (ICT) and direct and indirect network effects (Rochet & Tirole, 2006) as well as the growing importance of machine learning and AI technology. In recent years, there has been a significant acceleration in the development and employment of AI, with strong differ-

Note: The statements and views expressed in this chapter are those of the authors and do not necessarily reflect the views of the Federal Ministry for Economic Affairs and Climate Action.

Richard von Maydell, ETH Zürich, Zürich, e-mail: rbaron@ethz.ch
Christoph Menzel, Federal Ministry for Economic Affairs and Climate Action, Berlin,
e-mail: christoph.menzel@bmwk.bund.de

https://doi.org/10.1515/9783110790320-001

ences in the corporate integration of AI between countries and between sectors. The corporate development and employment of AI offer a wide range of new and often not yet foreseeable opportunities, from an increase in efficiency and sustainability in production (Industry 4.0) to targeted, customisable provision of products and services to entirely new business models. Nonetheless, there is an increasing debate about the extent to which AI will lead to imperfect competition, price markups and market barriers – as it has happened in the Tech Giant-dominated software industry with the rise of ICT. Especially online markets exhibit high profit margins and low rates of market entry hinting at significant market barriers (Stigler Committee, 2019). The increasing corporate integration of AI has the potential to further reinforce market concentration and create situations where leading companies ("first movers") gain a dominant position in the market.

As an appropriate dataset that would allow for an analysis of the effect of corporate AI integration on market concentration is not yet available, we examine investments in intangible assets (e.g., software and databases) as an essential prerequisite for the deployment of AI. We investigate how market concentration is linked to markups and intangible investments and discuss policy interventions that may decrease the risk of declining competition in digital economies. We call upon the introduction of an early warning system for detecting competition-hampering trends in digital markets and modernisation of anti-trust regulation considering the increasing importance of AI.

1.2 Literature

Existing literature provides a distinction between "weak" and "strong" AI, the former describing the software-driven solution of specific application problems and the latter characterising a technology that can reproduce (at least) human intelligence (Menzel & Winkler, 2018). However, an overarching and general definition of AI has not yet been clearly established and we therefore focus on the following definition: AI describes algorithms and their use in software tools that simplify tasks such as search operations, pattern recognition, inference or planning ("weak AI"). In contrast to previous ICT technologies, however, AI is unique as it can develop autonomously through self-learning properties (Brynjolfsson, Rock & Syverson, 2017) by application and training, for example through deep machine learning or reinforcement learning (Lu, 2020) and can therefore be described as a self-learning technology (Gersbach, Komarov & Maydell, 2022). Furthermore, AI is described as an intangible asset that is easy to scale up since AI software can be used simultaneously by multiple users (in a non-rival fashion) at relatively low variable costs. With growing regularity, the topic of AI is considered separately in the economic literature, as distinct from automation (Aghion, Jones & Jones, 2017), and discussed in connection with topics such as the economics of data (Jones & Tonetti, 2020). For example, Trammell and Korinek (2020) address the effects of AI on the employment structure, income structure and resulting productivity effects. In addition to the classic economic factors of production – namely labour, capital and natural resources – data, software and AI are playing an increasingly important role in digital economies (Varian, 2018; Wagner, 2020). Schweitzer et al. (2022) refer to the necessity for investments in AI for its training and development and point out that the notion that only a few firms can afford to invest in data-driven AI tends to further increase the risk of monopolisation in concentrated markets as its control will be increasingly concentrated in the hands of AI providers (Wagner, 2020).

Akcigit et al. (2021) note that average market concentration at the macroeconomic level has increased, firms' markups and profitability have increased significantly and that market shares are increasingly concentrated in single companies with high markups, which is particularly pronounced in the ICT and pharmaceutical industries. Recent studies (Bajgar, Criscuolo & Timmis, 2021; Goldin et al., 2020; de Loecker, Eeckhout & Unger, 2020; de Ridder, 2019) show that increasing investments in intangible assets have been a key driver for the rise in market concentration, markups and corporate profits over the last two decades.

Increased market concentration does not necessarily imply weakened competition but can also be an indicator of allocative efficiency if innovative and productive firms have a high market share (Bajgar, Criscuolo & Timmis, 2021; Bighelli et al., 2021). However, Autor et al. (2020) and Raurich, Sala and Sorolla (2012) show that higher markups can be charged in markets with imperfect competition. Therefore, Effenberger et al. (2020) summarise that increased market concentration can be a sign of allocative efficiency but also of reduced competitive pressure and associated productivity losses. Davis and Orhangazi (2021) provide empirical evidence that market concentration has increased, especially in the retail and ICT sectors, whereas Calligaris, Criscuolo and Marcolin (2018) note that this leads to a decrease in firm dynamics. Bajgar, Criscuolo and Timmis (2021) show that an increase in market concentration occurred, especially in globalised and digitally-intensive industries. In this sense, Aghion et al. (2019) suppose that rising corporate concentration and higher profits have emerged due to advances in ICT. In this vein, Diez, Fan and Villegas-Sánchez (2021) observe such a rise in market concentration primarily in the information and communications industry. At the same time, Autor et al. (2020) speak of a rise of superstar firms, as mainly large firms with a high degree of ICT can become more productive and firms with high investments in intangible assets have higher profit margins and lower labour shares.

The economies of scale and scope in digital economies may constitute a great comparative advantage for dominant firms and impede contestability of firms with market dominance (Schweitzer et al., 2022). With the increasing importance of intangible assets, fixed costs incurred for the development of databases, or the acquisition of software are playing a greater role in the integration of new technologies into production processes (European Commission, 2021; Grossman & Oberfield, 2021). High fixed costs can have a significant impact on market competition when serving as market barriers for competing firms (Haskel & Westlake, 2017), as it is already visible today in the software industry.

1.3 Investments in Software and Data, and Market Concentration

A detailed dataset on corporate investments in AI applications and its effect on market concentration is not yet available. However, the IBM AI Adoption Index 2022, which is based on company surveys, provides initial indications of application areas, driving forces and barriers to AI by industry and country. Only around 25% of the 7502 companies surveyed indicate that they have a holistic strategy on AI across the organisation (International Business Machines Corporation, 2022). In Germany, for instance, the IBM Global AI Adoption Index 2022 shows that across industries, above all a need to reduce costs, automate processes (e.g., advancing “Industry 4.0” (Brühl, 2015)) and a shortage of skilled workers are promoting increased AI use in companies.

Table 1.1: Expenditures on software and databases in % of sales of all German firms by industry. Source: ZEW (2020) – annual German innovation Panel.

	2018	2020
Software/Inform. Service /Telecomm.	6.88	4.47
Information and communication	5.59	3.81
Electric engineering	1.22	3.12
Legal/business corporate consulting	1.46	2.17
Financial services	1.43	1.82
Glass/ceramics/stoneware	0.28	0.28
Mining	0.31	0.26
Food/beverages/tobacco	0.15	0.23
Metal production	0.17	0.22
Mineral oil	0.05	0.10

Investments in software and data infrastructure represent a prerequisite for the implementation of AI applications. With the help of the annual German innovation panel from 2018 and 2020, which is provided by the Centre for European Economic Research (ZEW), further insights into industry-specific investments in software and data can be obtained. Table 1.1 shows expenditures on software and databases as a percentage of turnover for all enterprises for selected economic industries in Germany (according to Level 2 NACE Rev. 2 classification).

We show the values for the five industries with the highest and with the lowest investments in software and data. The highest investments in software and data are noticeable in ICT, electronics, as well as financial and consultancy services. The lowest investments can be observed in mining, metal production and the food and beverages industries. In general, great heterogeneity can be identified between industry-specific expenditures. For example, in 2020, the expenditures in the industry for software and ICT were 44 times higher than in the sector for mineral oil.

In the following, market concentration in different economic sectors is considered and placed in the context of investments in intangible assets. The analysis of sector-specific market concentration is based on the Herfindahl-Hirschmann Index (HHI).[1] The HHI for each aggregated economic sector (Level 1 by NACE Rev. 2 classification) is determined[2] and the development of the HHI in Germany per economic sector is illustrated in Figure 1.1. Based on the HHI, we notice strong differences in revenue concentration between economic sectors. In 2019, we observe the maximum market concentration (HHI = 2015) in the financial and insurance services sector, while the lowest value (HHI = 23) is in real estate and housing. We note that low values of the HHI are realised primarily in the economic sectors of real estate and housing, hotels and restaurants, construction and water supply. In contrast, we observe high values of the HHI, especially in the sectors of ICT, energy supply, mining, financial and insurance services.

Moreover, we observe strong differences in the fluctuations of market concentration between the economic sectors. While there are hardly any year-related differences in real estate and housing or water supply, there are large temporal variations in mining, ICT or energy supply, which could speak for a changing market dominance of firms but will not be further examined. To highlight the link between investments in intangible assets,[3] market concentration and markups, we depict several scatterplots in Figure 1.2, which are based on CompNet data for the countries France, Germany,

1 The HHI is determined using the following equation, where x_i represents the turnover of a single firm i and $H = \sum_{i=1}^{N} \left(\frac{x_i}{\sum_{j=1}^{N} x_j} \right)^2$. The value of the HHI obtained is multiplied by 10,000 and thus takes values between $\frac{10{,}000}{N} \leq H \leq 10{,}000$. The higher the HHI, the higher the market concentration, with values below 1000 considered harmless and above 1800 critical (Effenberger et al., 2020).

2 For the calculation, a weighted average of all economic classes is determined based on the total revenue in order to determine the HHI for an entire economic sector. Another measure for assessing market concentration is the concentration rate (CR), which we calculate for each economic sector based on official data from the Federal Statistical Office. Using this measure, we arrive at similar results regarding sector-specific concentration.

3 We report the mean of the total firm-specific real intangible investments in each 2-digit industry. This comprises all intangible investments and not only investments in software and data.

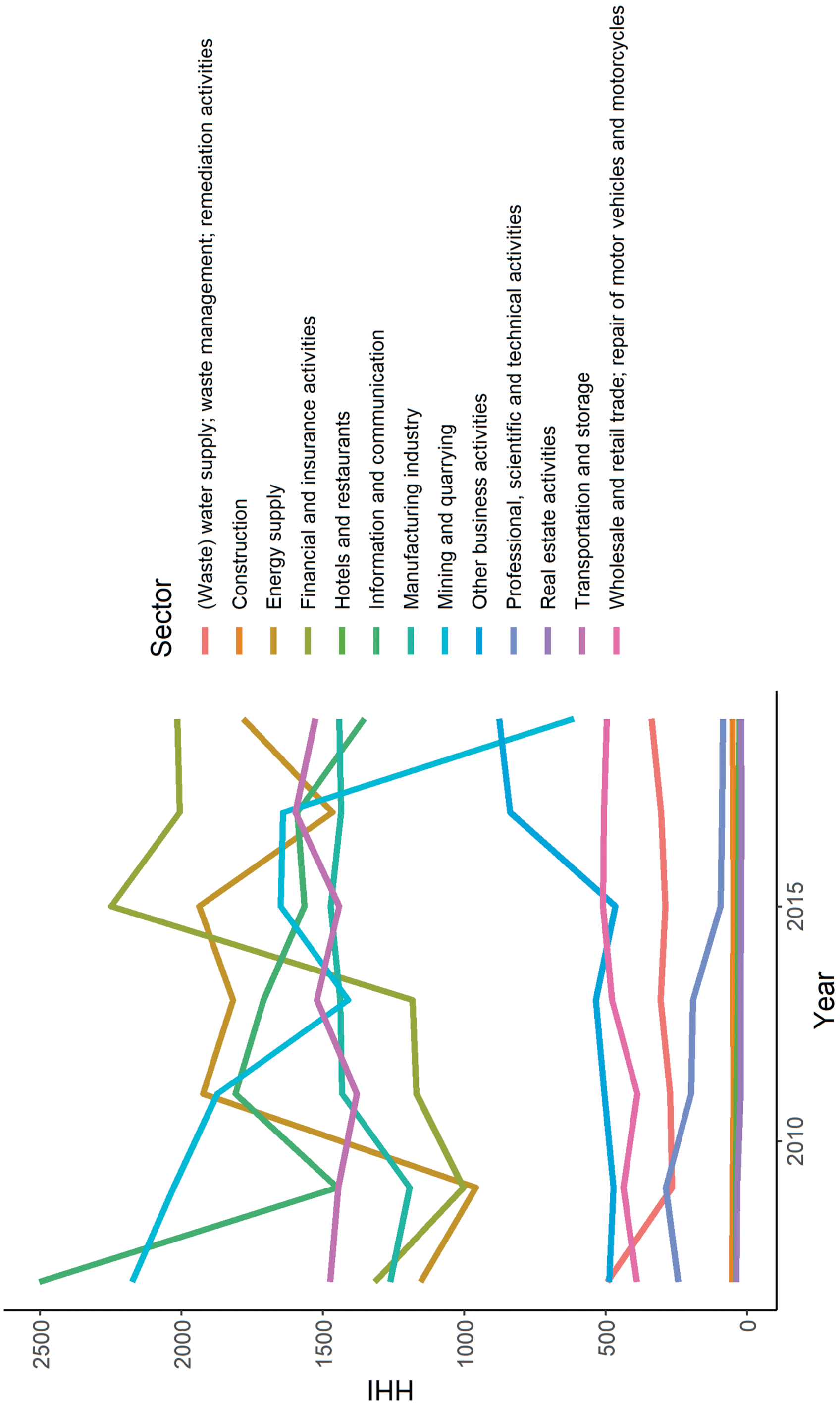


Figure 1.1: Market concentration measured by HHI per economic sector. Source: 8th Vintage CompNet dataset.

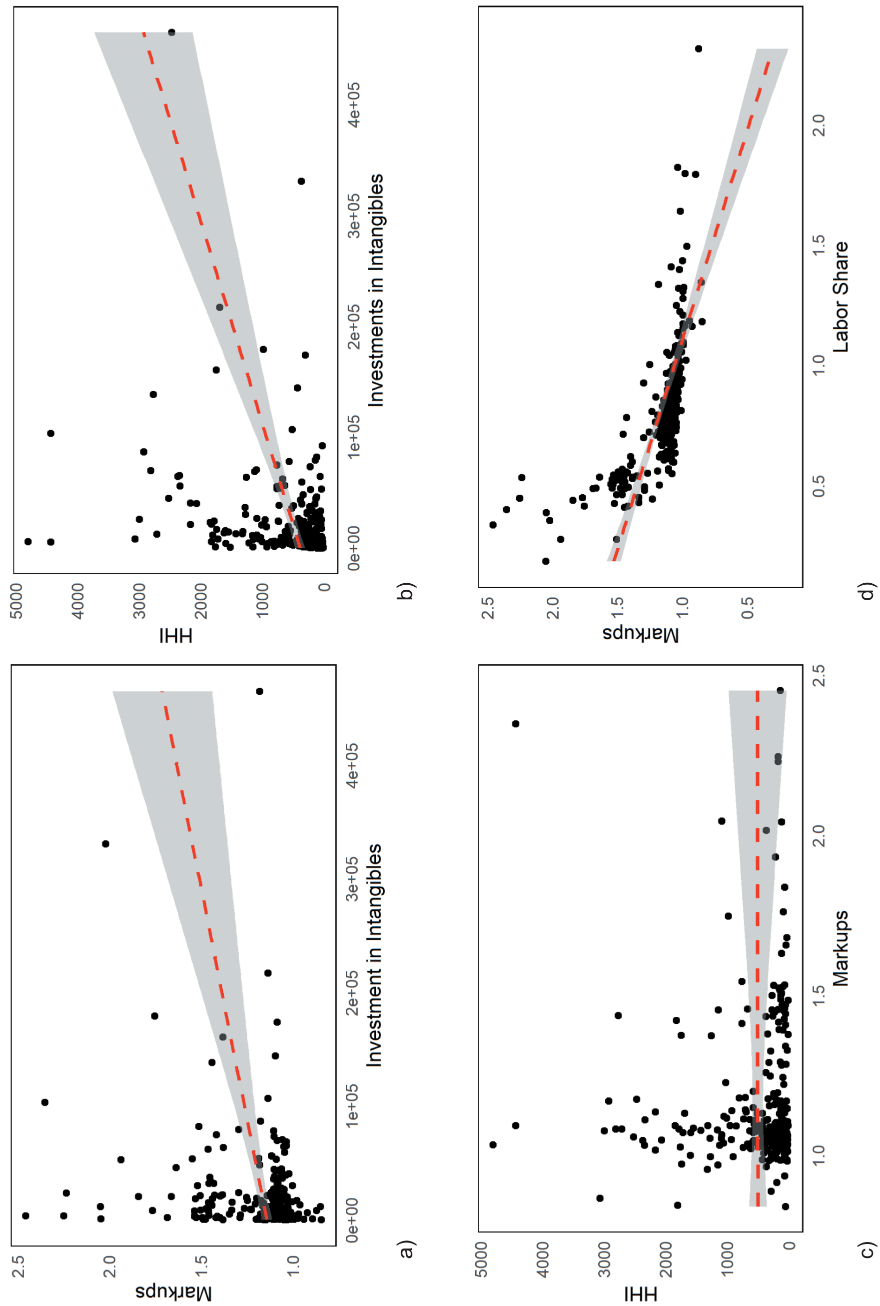


Figure 1.2: Scatterplots: Markups, investments in intangible assets, labour share and market concentration. Source: 8th Vintage CompNet dataset.

the Netherlands, Switzerland, Italy, Spain and Poland.[4] An observation point represents a specific economic industry (Level 2 by NACE Rev. 2 classification) from the abovementioned countries in 2018.[5]

We note that industries with high investments in intangible assets and industries with high market concentration show higher markups on average. The provided confidence intervals additionally indicate a positive relationship between markups, market concentration and investments in intangibles. Yet, no clear link between markups and market concentration can be observed. Moreover, it should be emphasised that markups are particularly associated with a lower labour share, which could be indicative of an increasingly unequal income distribution, where asset owners benefit from markups at the expense of the labour income. As mentioned in Section 1.2, existing literature points to the relevance of intangible assets for the increase in market concentration, especially in the ICT sector. A descriptive analysis of the available data, for Europe and especially Germany, provides fundamental support for this hypothesis. In Germany, high investment in software and data can be observed in the ICT-intensive sectors (see Table 1.1), as well as high market concentration in the energy, ICT, financial and insurance services sectors (see Figure 1.1).

Based on the available data, we can draw initial conclusions about the interplay between investments in intangible assets, market concentration, markups and the labour share, along with potential implications for competition. In Germany, we observe substantial investments in software and data, particularly within sectors already characterised by high market concentration, such as ICT and financial and insurance services. We posit that investments in AI could potentially exacerbate market concentration, leading to elevated price markups and diminished labour shares.

Nonetheless, the relation between AI and the competition indicators under investigation remains unclear. On the one hand, new and innovative companies could use AI to challenge the market dominance of individual companies in the future. On the other hand, in sectors in which high market concentrations have already been observed in recent years, such as in the energy, financial and insurance services or ICT sector, a stronger corporate integration of AI could reduce competitiveness, potentially leading to allocative inefficiency. However, more detailed empirical research, appropriate identification methods and new data with more information on the economic integration of AI are needed to assess the (causal) relationship between AI and market competition. Nonetheless, we subsequently delve into potential policy interventions aimed at mitigating the risks of AI-induced competition distortions.

---

4 CompNet is the Competitiveness Research Network, which provides micro-aggregated indicators computed by national data providers using firm-level data.

5 We remove all observations where no information on investments on intangibles are provided. We report 356 observations on 56 different industries. Moreover, we calculate 95% confidence intervals for our linear smoothing method using a t-distribution. Yet, our graphical illustration does not allow for assessing country- or industry-specific differences.

1.4 Policy Implications

The Tech Giants have built up their powerful market position in online services such as search engines, social networks or mediation, mainly due to the rising importance of ICT. We start our discussion of policy implications based on the premise that similar developments can also be expected due to the ascent of AI. At the EU level, an initial step for regulating digital markets was made with the Digital Market Act focusing on central platform services and gatekeeping firms with large market shares. In Germany, the GWB was mainly a result of the recommendations of the Monopolkommission (2018), promoting an adaptation of competition policy and legal framework to the digital transformation for counteracting the resulting trends of reduced firm competition in digital markets. Yet, the effectiveness of newly established regulatory interventions, considering the increasing importance of AI, remains unclear. Increasingly, the European Union is considering far-reaching legislation on AI for fundamental rights and safety requirements of AI systems (European Commission, 2021b). Yet, we stress the need for a thorough evaluation of potential competition challenges in markets that are becoming progressively AI-intensive.

1.4.1 Challenges in Regulating Digital Economies

Podszun (2022) refers to the fundamental architecture of competition law and the enforcement of classical anti-trust law, which increasingly shows significant inefficiencies in digital ecosystems. More specifically, Furman and Seamans (2019) and Wagner (2020) refer to the risk of winner-take-all market structures that could make competition more complicated in AI-intensive industries. Especially in concentrated digital markets, firms have the potential to harm the competitiveness as monopolistic firms can personalise decisions to users and firms (Hardt, Jagadeesan & Mendler-Dünner, 2022) and can thus steer consumers towards more profitable behaviour. Existing antitrust concepts have struggle in identifying anti-competitive patterns in digital ecosystems (Hardt, Jagadeesan & Mendler-Dünner, 2022), especially regarding platform competition in digital markets. The reason is that several competition measures (i.e., Lerner or Herfindahl-Hirschmann Index) need appropriate market definitions to be able to detect market dominance. Due to the difficulty that market boundaries can only be hardly drawn (Stigler Committee, 2019) in digital economies, it is difficult to employ appropriate measures for an effective competition enforcement in case of market failure in digital ecosystems. Thus, the definition of market boundaries needs to be adjusted to new developments in digital economies as a step towards modernising competition law.

In addition to an appropriate taxation of digital firms and an internationally consistent reform of the tax system (Faulhaber, 2019), novel intervention options should be discussed in order to guarantee a market order that promotes competition in software-intensive industries. To stimulate growth-enhancing AI competition and share the benefits of AI with all actors in an economy, a tax on profits in digital markets with imperfect competition or a so-called “AI tax” (Gersbach, Komarov & Maydell, 2022) could be introduced. However, further scientific research on such tax policy interventions is needed to adequately assess the effectiveness of such measures. Nevertheless, the system of privatised gains and socialised losses in digital economies should be further reviewed and modified, i.e., by the introduction of an international minimum tax (OECD, 2021).

Furthermore, new international legal frameworks and patent regulations are needed to define intellectual property rights for non-human work, in particular for software and AI (Ernst, Merola & Samaan, 2019). A modernisation of property rights to be agile and reflexive would be advantageous to enable a co-ownership of data users and data holders that does not disproportionately favour any of the parties (Schweitzer et al., 2022) and give each party the independent right to use the data without the approval of other parties.

1.4.2 Measures Supporting the Economic Integration of AI

The aim should be to create incentives for AI development, to promote the availability of data and AI applications as a public good and to prevent possible – sometimes “natural” (network and scale effects) – monopolisation tendencies. Diez, Fan and Villegas-Sánchez (2021) argue that especially the elimination of obstacles that hinder technologically lagging firms (referring to their potential for corporate AI integration) from competing with a technological frontier is an appropriate policy intervention. Therefore, we propose that technologically lagging firms should be subsidised to invest in AI infrastructure to be better able to compete against technologically more advanced firms. One idea, for example, might be to grant technologically lagging firms with funding for their expansion of AI infrastructure (i.e. servers, data centres) or their general AI productivity in order to promote their competitiveness. In addition, access to data, software and the use of AI should be increasingly promoted to support innovative, small and medium-sized firms in the research, development and deployment of AI. Yet, the IBM Global AI Adoption Index 2022 points to the current challenges in corporate AI integration that persist despite AI support measures such as the AI/Blockchain Investment Support Programme at the European level (European Commission, 2021). The absence of AI know-how continues to pose a significant barrier to the integration of AI in businesses. This is succeeded by challenges in effectively integrating and scaling AI projects, dealing with high AI expenses, and the lack of a comprehensive corporate strategy for AI implementation (International Business Machines Corporation, 2022).

Schweitzer et al. (2022) point out that only a few firms nowadays consider data sharing as a relevant feature of their operating business model and assign only minor importance to the acquisition of external data. There is little empirical research on the importance of data sharing, data portability and interoperability which is characterised by issues related to liability, privacy, data and cyber security. Nonetheless, interoperable standards for data sharing should be established and trust and communication in digital economies should be promoted, especially in the presence of substantial network externalities (Stigler Committee, 2019). In this vein, Wagner (2020) emphasises that legislature and property rights should facilitate data use, provision and sharing.

1.4.3 Early Warning System and Modernisation of Competition Policy

From a competition policy perspective, it is central to recognise increasing concentrations and decreasing competitive pressure on digitally-intensive markets and to be able to counteract anti-competitive developments at an early stage. There is a need for anti-trust interventions to regulate data-driven monopolies at an early stage and not only after realising that firms have reached market dominance. Pro-competitive measures should be on the political agenda in times of a growing AI integration and increasing income inequality in data-based economies. Yet, the question arises at which point regulatory institutions should intervene to not suffocate innovation but still foster competition. Enforcement practices need to be adapted to be able to cope with new challenges in digital economies. We argue that there is an increasing necessity for modernised anti-trust regulation and enforcement of a new legislation to prohibit superstar firms from impeding market entry of start-ups, leading to less competition and innovation. Furthermore, smart legislation with *competition by design* (Podszun, 2022) should be promoted.

There is a need for anti-trust intervention options that can regulate abusive market power in software-intensive industries at an early stage and not only ex-post in the case of classic market dominance, in order to be able to guarantee a pro-competitive market order. Current regulatory measures such as the DMA, Data Governance Act or American Innovation and Choice Online Act have the weakness that they can hardly react flexibly to new technological developments and resulting distortions of competition and abuse of market power. New competition policy instruments need to be developed or established instruments, such as merger control, must be modernised to be more effective.

An ideal concept would be the establishment of an early-warning system enabling market authorities to distinguish between harmful and monopolistic competition and solely highly concentrated markets where innovative activities are prevailed. Experts in different fields should consider the possibilities of improving ex-ante regulations that foster competition and enable the rise of innovative firms. For example, a multi-factor indicator could be developed such that an independent competition authority can detect imperfect competitive markets and intervene, if necessary. A possibility to modernise anti-trust would be a transition from an adversarial to a cooperative enforcement in increasingly digitally-intensive markets. Yet, such novel political measures should also be practically put to the test for example, in real-world laboratories and regulatory sandboxes – to answer current competition-related questions in increasingly AI-intensive markets. Additionally, data intermediaries could be more intensively involved for assisting in merger controls and competition regulation for hampering collusion of market-dominating firms.

1.5 Conclusion

AI is already considered one of the key technologies that will determine the competitiveness of economies in an increasingly digitalised world (European Commission, 2021). A detailed report of investments in AI applications in companies combined with information on market concentration is not yet available as a data set. Yet, investments in intangible investments (e.g., computer programs, databases and copyrights) represent a prerequisite for the implementation of AI applications. In Germany, for example, high expenditures on software and data can be particularly noticed in the ICT, electronics and financial and insurance services sectors. With the help of descriptive analyses based on data from the annual German innovation survey and information from the CompNet database, we identify increased price markups and high market concentration in industries with high investment in intangible assets in the European region. In view of the increasing importance of AI, we emphasise that economic research should examine more profoundly to what extent an increased use of AI applications poses an additional risk to competition, especially in sectors where high market concentration and markups already prevail.

There is still uncertainty about the legal development and the scope of regulatory measures in the digital economy, which may further complicate the integration of data, software and AI into business processes. New private law concepts, anti-trust regulation and contemporary merger control laws are needed to enable an adequate balance and flexibility between regulation and shaping of the market that enables market integration of data-driven applications and reduces monopoly risks in digital economies (Schweitzer et al., 2022). A consistent international legal framework for the application of AI must first be developed and overlapping and interlocking sets of rules need to be examined. Furthermore, there is a need for financial support to facilitate corporate AI integration, serving both as a catalyst for technological advancement and a tool to foster competition. Finally, we advocate for early-stage adjustments to competition and anti-trust policies to prevent allocative inefficiencies in rapidly digitising economies.

Bibliography

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., & Li, H. (2019). A Theory of Falling Growth and Rising Rents. NBER Working Paper, No. 26338.

Aghion, P., Jones, B. F., & Jones, C. I. (2017). Artificial Intelligence and Economic Growth. NBER Working Paper, No. 23928.

Akcigit, U., Chen, W., Diez, F. J., Duval, R. A., Engler, P., Fan, J., Maggi, C., Tavares, M. M., Schwarz, D. A., Shibata, I., & Villegas-Sánchez, C. (2021). Rising Corporate Market Power: Emerging Policy Issues. IMF Staff Discussion Note, SDN/21/01.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., & van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *Quarterly Journal of Economics*, 135(2), 645–709.

Bajgar, M., Criscuolo, C., & Timmis, J. (2021). Intangibles and industry concentration. OECD Science, Technology and Industry Working Papers, OECD Publishing, Paris, 2021(12). https://doi.org/10.1787/ce813aa5-en.

Bighelli, T., di Mauro, F., Melitz, M. J., & Mertens, M. (2021). European Firm Concentration and Aggregate Productivity. IWH-CompNet Discussion Papers, No 3.2021.

Brühl, V. (2015). Wirtschaft des 21. Jahrhunderts. Herausforderungen in der Hightech-Ökonomie. Wiesbaden: Springer Gabler. https://doi.org/10.1007/978-3-658-04883-9.

Brynjolfsson, E., Rock, D., & Syverson, C. (2017). Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics. NBER Working Paper, No. 24001.

Bundesregierung. (2022). Digital-Gipfel 2022: Gemeinsam Innovationen durch Daten ermöglichen. https://bmdv.bund.de/SharedDocs/DE/Pressemitteilungen/2022/094-wissing-innovation-durch-daten.html (accessed 29 August 2023).

Calligaris, S., Criscuolo, C., & Marcolin, L. (2018). Mark-ups in the Digital Era. OECD Science, Technology and Industry Working Papers, 2018(10).

Davis, L., & Orhangazi, Ö. (2021). Competition and monopoly in the US economy: What do the industrial concentration data show? *Competition & Change*, 25(1), 3–30.

de Loecker, J., Eeckhout, J., & Unger, G. (2020). The rise of market power and the macroeconomic implications. *Quarterly Journal of Economics*, 135(2), 561–644.

de Ridder, M. (2019). Market power and innovation in the intangible economy (Issue No. 1931). https://ideas.repec.org/p/cam/camdae/1931.html.

Diez, M. F. J., Fan, J., & Villegas-Sánchez, C. (2021). Global declining competition. *Journal of International Economics*, 132(103492). https://doi.org/10.1016/j.jinteco.2021.103492.

Effenberger, A., Enkelmann, S., Menzel, C., Neumann, D., & Stolle, J. (2020). Marktkonzentration, Produktivität und Preisaufschläge: Deskriptive Evidenz auf Basis amtlicher Daten für Deutschland. Bundesministerium Für Wirtschaft Und Energie Discussion Paper, No. 9.

Ernst, E., Merola, R., & Samaan, D. (2019). Economics of artificial intelligence: implications for the future of work. *IZA Journal of Labor Policy*, 9(1).

European Commission. (2021a). Fostering a European Approach to Artificial Intelligence. Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, 205 final.

European Commission. (2021b). Laying down harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts. Proposal for a Regulation of the European Parliament and of the Council. https://ec.europa.eu/info/law/better-regulation/ (accessed 28 September 2021).

Faulhaber, L. v. (2019). Taxing tech: The future of digital taxation. Virginia Tax Review, 39, 145.

Furman, J., & Seamans, R. (2019). AI and the economy. *Innovation Policy and the Economy*, 19(1), 161–191.

Gersbach, H., Komarov, E., & von Maydell, R. (2022). Artificial Intelligence as Self-Learning Capital. CEPR Discussion Paper, No. 17221.

Goldin, I., Koutroumpis, P., Lafond, F., & Winkler, J. (2020). Why Is Productivity Slowing Down? Munich Personal RePEc Archive Working Paper, No. 107644. https://mpra.ub.uni-muenchen.de/107644/.

Grossman, G. M., & Oberfield, E. (2021). The Elusive Explanation for the Declining Labor Share. NBER Working Paper, 29165. https://doi.org/10.3386/w29165.

Hardt, M., Jagadeesan, M., & Mendler-Dünner, C. (2022). Performative power. ArXiv Preprint ArXiv:2203.17232.

Haskel, J., & Westlake, S. (2017). *Capitalism Without Capital*. Princeton University Press.

International Business Machines Corporation. (2022). IBM Global AI Adoption Index. Armonk, NY: IBM Corporation, 2022. Report.

Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review*, 110(9), 2819–2858.

Lu, C.-H. (2020). Artificial intelligence and human jobs. *Macroeconomic Dynamics*, 25(5), 1–40.

Menzel, C., & Winkler, C. (2018). Zur Diskussion der Effekte Künstlicher Intelligenz in der wirtschaftswissenschaftlichen Literatur. Bundesministerium Für Wirtschaft Und Energie Discussion Paper, No. 8.

Monopolkommission, W. (2018). XXII. Hauptgutachten Der Monopolkommission, § 44 Abs. 1 Satz 1 GWB, 3. Juli 2018, 44.

OECD. (2021). Tax Challenges Arising from Digitalisation of the Economy – Global Anti-Base Erosion Model Rules (Pillar Two). OECD. https://doi.org/10.1787/782bac33-en.

Podszun, R. (2022). Competition in the Digital Economy: What next after the Digital Markets Act? – Statement for the Economic Committee of the German Parliament (Bundestag). Statement for the Economic Committee of the German Parliament (Bundestag) (27 April 2022).

Raurich, X., Sala, H., & Sorolla, V. (2012). Factor shares, the price markup, and the elasticity of substitution between capital and labor. *Journal of Macroeconomics*, 34(1), 181–198.

Rochet, J.-C., & Tirole, J. (2006). Two-sided markets: A progress report. *The RAND Journal of Economics*, 37(3), 645–667.

Schweitzer, H., Metzger, A., Blind, K., Richter, H., Niebel, C., & Gutmann, F. (2022). Data Access and Sharing in Germany and in the EU: Towards a Coherent Legal Framework for the Emerging Data Economy: A Legal, Economic and Competition Policy Angle-Final Report. Report on Behalf of the German Ministry for Economic Affairs and Climate Change.

Stigler Committee. (2019). *Stigler Committee Final Report on Digital Platforms*. The University of Chicago Booth School of Business.

Trammell, P., & Korinek, A. (2020). Economic Growth under Transformative AI. GPI Working Paper, 2020(8).

Varian, H. (2018). Artificial intelligence, economics, and industrial organization. In: *The Economics of Artificial Intelligence: An Agenda* (Ajay Agrawal, Joshua Gans, and Avi Goldfarb) (pp. 399–419). Chicago: University of Chicago Press.

Wagner, D. N. (2020). Economic patterns in a world with artificial intelligence. *Evolutionary and Institutional Economics Review*, 17(1), 111–131.

Part 2: **Production and Processes**

Markus Feld, Wolfgang Arens-Fischer, Marcel Schumacher

Chapter 2
"KI-AGIL" – An Agile Process Model to Make AI Development Accessible to SMEs

Abstract: Artificial Intelligence (AI) is considered a key technology of digital transformation. Most companies agree that machine learning offers many opportunities for improving their own products and services or even developing new business models. However, small and medium-sized companies (SMEs) often lack the expertise to develop their own AI solutions and are therefore unable to gain access to these opportunities.

The INTERREG research project KI-AGIL had the goal of supporting SMEs in the German–Dutch border region in developing their own AI solutions. As part of the project, the University of Applied Sciences Osnabrück developed and field-tested an agile process model that specifically addresses the needs and prerequisites of SMEs that have little or no previous experience in AI development. It became evident that especially the explorative approach allows companies a low-threshold access to the development of their own AI solutions.

Keywords: agile process model, AI development, machine learning, SME

2.1 Introduction

Artificial intelligence (AI) is considered a key technology of the digital transformation. Enormous progress has been made in recent years, especially in the field of machine learning. However, many companies lack the expertise to develop AI-supported technologies. In particular, small and medium-sized enterprises (SMEs) often do not yet know how to harness this potential for themselves. In an expert survey in 2021, a lack of expertise and skilled workers was identified as the primary obstacle for the use of AI in SMEs. Approximately 75% of the experts rated this as a very strong obstacle and around 20% considered it a strong obstacle (Begleitforschung Mittelstand Digital, 2021). A Deloitte study from 2020 came to similar conclusions: Out of more than 300 German SMEs surveyed, the lack of expertise was rated as the significantly greatest obstacle to AI in SMEs, with 65% agreeing (Deloitte, 2021).

Markus Feld, Institute for Dual Study Programs, University of Applied Sciences Osnabrück – Campus Lingen, e-mail: m.feld@hs-osnabrueck.de
Wolfgang Arens-Fischer, Institute for Dual Study Programs, University of Applied Sciences Osnabrück – Campus Lingen, e-mail: w.arens-fischer@hs-osnabrueck.de
Marcel Schumacher, Institute for Management and Technology, University of Applied Sciences Osnabrück – Campus Lingen, e-mail: marcel.schumacher@hs-osnabrueck.de

https://doi.org/10.1515/9783110790320-002

The German–Dutch INTERREG research and development project KI-AGIL addresses exactly this problem. The KI-AGIL project started in July 2020 and ended in December 2022. The goal of the project was to open up the development of modern AI technologies for regional companies. The focus was deliberately on companies that had little to no previous experience in developing their own AI solutions. The focus was set on the following research questions:

- How can agile ways of working support SMEs in developing their own AI solutions?
- How can the findings from the project be transferred into a standardized innovation process that SMEs can use to develop and use AI solutions?

Therefore, six selected SMEs in the German–Dutch border region were supported by the University of Applied Sciences Osnabrück and the Hanze University of Applied Sciences to identify individual use cases for implementing AI in their production, products and services to develop corresponding AI technologies. Agile work methods were employed during the development of the corresponding AI technologies. The agile approach was tested and its findings were used to create an agile process model that can be adapted by other SMEs.

2.2 Related Work

Agile process models have been established in software development for many years and have proven to be an effective method for solving complex problems (Vogel, 2021). While classic, sequential process models require a precise definition of the requirements at the beginning of the project, agile process models enable an iterative approach to problem-solving. Agile approaches can thus facilitate entry into new business areas and technologies by reducing complexity at the beginning and enabling an explorative approach.

Although originally developed as a process model for data mining, today the Cross-Industry Standard Process for Data Mining (CRISP-DM) is often used in AI development (Chapman et al., 2000; Schröer, Kruse & Gómez, 2021). Under the name CRISP-ML(Q) (CRoss-Industry Standard Process model for the development of Machine Learning applications with Quality assurance methodology), a further development of the model was presented in 2021 (Studer et al., 2021). The aim here was to adapt the model more closely to the development of machine learning models and to add additional aspects relating to quality assurance.

Both models are very similar in their basic structure. Table 2.1 shows a comparison of the phases of the two models. As can be seen, the main difference in the structure is that in CRISP-ML(Q) the phases “business understanding” and “data understanding” were combined into one phase in order to emphasise the close interconnection between the activities of these phases. While a cycle of the CRISP-DM model ends with the

Table 2.1: Comparison of the phases of the process models CRISP-DM and CRISP-ML(Q) (Studer et al., 2021).

CRISP-DM	CRISP-ML(Q)
Business understanding	Business understanding and data understanding
Data understanding	
Data preparation	Data preparation
Modelling	Modelling
Evaluation	Evaluation
Deployment	Deployment
–/–	Monitoring and maintenance

deployment phase, CRISP-ML(Q) adds an additional phase “monitoring and maintenance.” In this phase, the permanent operation of the application is to be regulated and monitored. One of the aims is to ensure that the prediction quality of the model does not degrade in the long term (Studer et al., 2021). The cause of such degradation could be, for example, a creeping change in input data due to changing external conditions.

Both models focus very much on the development process of a single AI solution. CRISP-ML(Q) complements this by considering the complete lifecycle of the developed solution. Both models already assume a certain understanding of the use case to get started. However, this does not necessarily apply to companies that are dealing with the development of their own AI solutions for the first time.

Another similarity between the two models is that only the solution that is specifically developed in each respective cycle is considered. Both models offer the possibility of building on the results of previous phases when planning a new cycle and defining new use cases through the resulting better understanding. However, what is missing is a higher planning level that enables the definition of long- and medium-term goals. Such goals are essential to align the individual cycles with these goals and thus to be able to implement an iterative solution for complex problems. Such mechanisms, on the other hand, are an integral part of most agile process models. In Scrum, one of the most widespread models in agile software development, this mechanism is implemented through the product backlog. In the product backlog, corresponding goals can be defined at a very high level of abstraction. In the Scrum Guide, published by the developers of Scrum, this is referred to as the product goal (Schwaber & Sutherland, 2020).

To ideally support an iterative approach in the development of AI solutions, it is therefore useful to develop a process model that enables both the higher planning level and is tailored to the specificities of AI development, thus combining the best of the above-mentioned models.

2.3 Process Model for Agile AI Development

As part of the project KI-AGIL, a process model was developed that is specifically designed to enable SMEs to easily get started with developing their own AI solutions. The aim is on the one hand to keep entry barriers as low as possible and on the other hand to quickly incorporate the knowledge and results that arise during project implementation into further development.

The developed process model is presented in Figure 2.1. The individual phases of the model will be presented separately below.

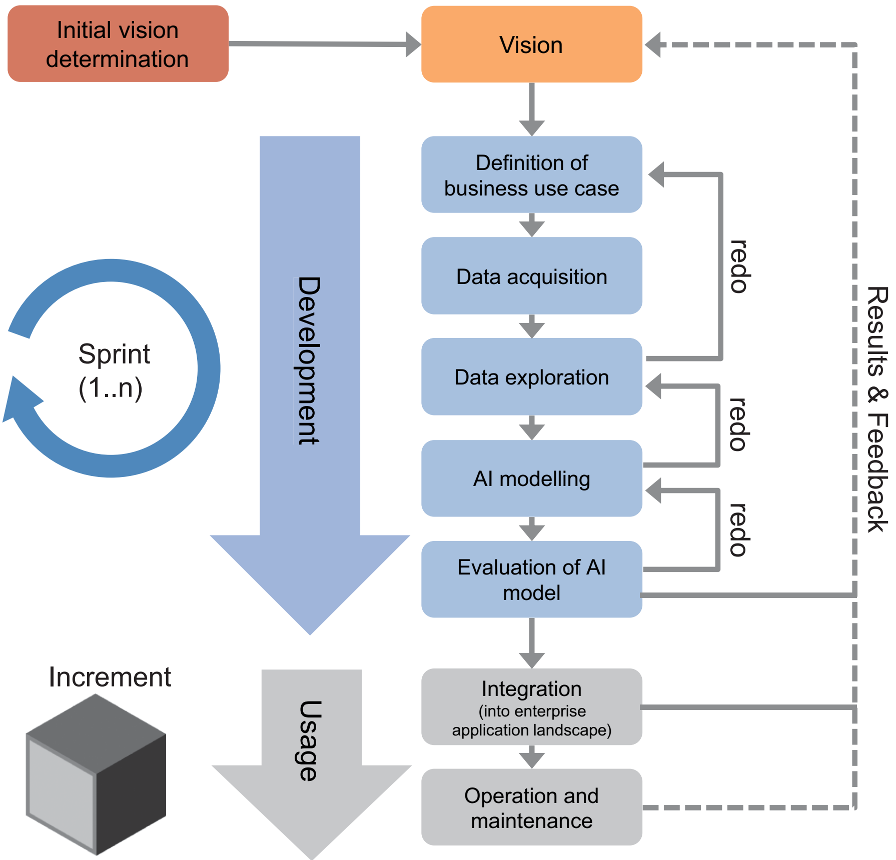


Figure 2.1: KI-AGIL process model for agile AI development. (compiled by authors.)

2.3.1 Initial Vision Determination and Vision Statement

Existing process models such as the CRISP-DM usually start directly with the implementation of a concrete use case (see Section 2.2). In practice, it has been shown that companies with little or no experience in AI development are often not yet able to describe and delineate potential use cases in detail right at the beginning of the project. For this reason, the KI-AGIL process model initially provides a separate phase for a preceding investigation before the actual project starts. This phase serves to identify potential fields of application and to evaluate them regarding their feasibility and the expected benefits. As a result of this phase, an application field is selected and a suitable vision statement is defined. The vision statement describes what is to be achieved in the medium to long-term perspective through the use of AI. The vision statement thus sets the direction of development and serves as a point of orientation for further development in all subsequent steps.

The primary focus here is not on achieving short- or medium-term technological or economic goals. Rather, this often has a strong influence on the strategic orientation of the company itself. In the KI-AGIL project, intensive reflection processes took place between the company management and the university in this context. In the course of the project, the increasing use of AI can generate further learning and improvement processes, which can lead to changes in the vision and desired customer benefits of the business models.

2.3.2 Development Phase (Sprint)

Starting from the defined vision statement, the agile development of the AI solution begins in successive sprints. Each sprint has the structure described below.

2.3.2.1 Definition of the Business Use Case

Each sprint starts with the definition of the use case to be implemented. The use case must originate from the previously defined application areas (see Section 2.3.1) and be clearly assignable to the vision statement. Ideally, the first sprint should result in a usable (partial) solution in the form of a Minimum Viable Product (MVP). In this respect, it is of course important that the selected use case can be realised with the capacities and skills that are available within the sprint.

In the subsequent sprints, it is possible to successively expand the use case from the previous sprints or, alternatively, to define a new use case that also serves to achieve the vision statement.

2.3.2.2 Data Acquisition

In many process models, such as CRISP-DM or CRISP-ML(Q), the definition of the use case is directly followed by the data analysis (see e.g., the "data understanding" phase in CRISP-DM). The practical application of the model within the scope of KI-AGIL showed that the process of data acquisition can be quite challenging for smaller and medium-sized companies. In addition, this step often involves different employees compared to the subsequent data analysis.

To take these points into account and to make the effort for the companies transparent and thus easier to plan, a separate phase for data acquisition was added to the process model.

2.3.2.3 Data Exploration

After data acquisition has been completed, the data analysis step can start. This phase involves analysing the data provided in terms of data quality, cleaning the data and analysing and interpreting correlations in the data. In the course of the project, it became clear that close collaboration between the business experts and the data analysts is of crucial importance in this phase. In many cases, the targeted analysis of the data already led to new insights and a better understanding of the use case itself. In some cases, this improved understanding may lead to a reasonable return to the definition phase in order to refine the defined use case. On the other hand, if the data analysis reveals that the available data is not sufficient for the defined use case, a return to the definition phase must also be made. There, it must be checked whether the use case can be reasonably adapted based on the existing data or whether additional data sources can be found. If this is not possible, the sprint has to be aborted and a new use case needs to be selected. If the data analysis is successful, the transition to AI modelling takes place.

2.3.2.4 AI Modelling

In the AI modelling phase, the first step is to identify suitable AI methods. The decisive factors here are the problem to be solved and the available data. After the selection of the AI method, the AI model is created. This is immediately followed by the training and evaluation process in the next phase.

If no suitable model can be determined due to the prerequisites or if insufficient data is available for the models in question, it is possible to return to the data exploration phase.

2.3.2.5 Evaluation

In this step, the training of the model created in the previous AI modelling phase and the performance evaluation takes place. Based on this, the models are successively optimised (e.g., in the context of hyperparameter optimisation, adapted training strategies, etc.). When evaluating the models, in addition to the usual metrics for assessing prediction quality, methods from the field of Explainable AI should be used to ensure that the model makes decisions based on the right foundations. Making the decision-making process of AI transparent fulfils another important purpose: it creates trust in the AI and thus promotes acceptance (Theis et al., 2023). This is a crucial factor for AI solutions to be permanently established in the company, even beyond the project duration and without the support of the university.

If it becomes apparent during the evaluation that adjustments need to be made to the type or basic structure of the model, a return to the previous step of AI modelling can be made at any time. Especially in the development of more complex AI models, it will usually be necessary to go through the two steps of AI modelling and evaluation several times.

The development sprint ends with the successful evaluation. The results and findings from the completed sprint should be analysed and used to review and, if necessary, sharpen the vision statement. Based on this, the next sprint can be initialised.

Usage: If the result of the sprint represents an artefact that can be used independently, the model can be put into operation in parallel with the initialisation of the next development sprint.

2.3.2.6 Integration

For this purpose, integration into the productive IT landscape must at first be executed. In order for the AI model to be able to access the most up-to-date data, the necessary connections to the productive systems must be established. Depending on the use case, these can be internal and external databases, Manufacturing Execution Systems (MES), Enterprise Resource Planning (ERP) systems, etc. The training of the employees to use AI is also part of this step. An extensive testing phase should be performed before the actual go-live.

2.3.2.7 Operation and Maintenance

Once integration and testing have been successfully completed, going live can be executed. This marks the beginning of the actual operation of the model. Even during operation, the AI and its results should be continuously monitored. Early detection of errors or a decrease in the quality of AI results is essential to implement appropriate corrective measures, such as retraining the system.

2.3.3 Development Process Across Multiple Sprints

The development of more complex AI solutions takes place in a sequence and interlocking of several consecutive development prints. As stated before, the entire project starts with the initial vision determination phase, whose result is the overarching vision. As explained in the previous sections, this vision can initially be formulated in quite general terms. In each development sprint, a specific use case is defined and implemented to approach the vision. If the sprint produces a self-contained usable artefact, it can be used productively in the subsequent usage phase. At the same time, the next development phase can already begin.

Both the results and insights gained from the development phase and the ongoing operations can be used to review and, if necessary, refine or adjust the overarching vision. This gradual approach enables the solution of even more complex tasks.

Table 2.2 shows a comparison of the phases of the KI-AGIL process model with the two models CRISP-DM and CRISP-ML(Q). The following is a brief summary of the main differences:

To provide a low-threshold start for companies with little or no AI experience, the KI-AGIL process model starts with a preceding phase for the initial determination of the vision statement (for details, see Section 2.3.1). Furthermore, the KI-AGIL model introduces the additional phase of “data acquisition.” The phases “data understanding” and “data preparation,” on the other hand, are combined into the phase “data

Table 2.2: Comparison of the phases of CRISP-DM, CRISP-ML(Q) and the KI-AGIL process model.

<table>
<tr><th>CRISP-DM</th><th>CRISP-ML(Q)</th><th>KI-AGIL process model</th></tr>
<tr><td>–/–</td><td>–/–</td><td>Initial vision determination</td></tr>
<tr><td>Business understanding</td><td rowspan="3">Business understanding and data understanding</td><td>Definition of business use case</td></tr>
<tr><td></td><td>Data acquisition</td></tr>
<tr><td>Data understanding</td><td rowspan="2">Data exploration</td></tr>
<tr><td>Data preparation</td><td>Data preparation</td></tr>
<tr><td>Modelling</td><td>Modelling</td><td>AI modelling</td></tr>
<tr><td>Evaluation</td><td>Evaluation</td><td>Evaluation</td></tr>
<tr><td>Deployment</td><td>Deployment</td><td>Integration</td></tr>
<tr><td>–/–</td><td>Monitoring and maintenance</td><td>Operation and maintenance</td></tr>
</table>

exploration." This relates primarily to the different groups of experts involved in these phases in the companies (see Sections 2.3.2.2 and 2.3.2.3). The contents of the subsequent phases are comparable to the phases of the CRISP-ML(Q), besides the adapted wording to match the terms used in German–Dutch SMEs.

2.3.4 Development in Multidimensional Workspaces

Collaboration within the project occurred on multiple levels and in various forms of cooperation, which are referred to as workspaces. Within these workspaces, collaboration took place with different objectives. Through the integration and interaction of the individual workspaces, sustainable added value was created due to synergy effects.

2.3.4.1 Project-specific Workspaces

The central workspace for each sub-project within KI-AGIL was formed by the core project team, which consisted of employees from the company and the university. On the company side, the roles of business expert and data manager were essential. Since the participating companies did not have much experience with AI, the AI experts were provided by the university. If the company already had AI experts, they also took part in the project team. During the ongoing sprints, the interdisciplinary core project team met once a week at a fixed time to exchange information on the current status and to plan the next tasks. In addition, supplementary task-related working meetings were scheduled as needed.

The workspace of the core project team is concentrically enclosed and expanded by the workspace of the extended project team. In addition to the members of the core team, the extended project team included the company's management and representatives from other relevant departments such as sales and customer service. On the university side, the scientific management and programme managers complemented the extended project team. The extended project team met once a month. In this setting, the relevant intermediate results were presented and the further development direction was determined.

2.3.4.2 Inter-project Workspaces

Exchange between the sub-projects within KI-AGIL was promoted in several ways. The scientific management and programme management played an important role in this. As they were part of the extended project groups of the different sub-projects, they were able to provide a quick exchange of experience between the projects and, if necessary, establish a connection between the project participants of different sub-projects on a topic-related basis.

Another way in which the projects collaborated and exchanged knowledge was through the AI expert panels, which brought together all AI experts from the German and Dutch sides. The primary objective of these rounds was to facilitate the sharing of experiences and knowledge about the technologies, models, frameworks, etc. used in the projects.

2.3.4.3 Project-enclosing Workspace

The workspaces of the individual sub-projects (see Section 2.3.5.1) and the inter-project workspaces (see Section 2.3.5.2) are encompassed by the project-enclosing workspace, which also serves as the boundary of the project context. This workspace is used to explore overarching research questions, including the development and testing of the agile process model presented in this paper, alongside its corresponding role models. Additionally, investigations on topics such as technology acceptance are conducted in this workspace. Collaboration between the two participating research institutions, University of Applied Sciences Osnabrück and Hanze University of Applied Sciences, is a key focus at this level.

2.3.4.4 Project-external Workspaces

In addition to the project's internal collaboration, various external working spaces were also explored and utilised as part of the research project. This involved the exchange with external institutions, partners and networks to ensure that the project outcomes could benefit not only the participating companies but also other SMEs in the German–Dutch region.

The exchange in the business networks aims to reduce the obstacles SMEs may encounter in using AI. Mutual exchange makes it easier for companies to recognise the advantages of AI and to reflect them on their own business cases. In this process, the university can play an important role as a scientific partner and knowledge broker.

Companies are often faced with perceived risks when entering the field of AI development. These are largely due to the expected complexity of AI. For example, companies fear that they might not be able to develop an AI that delivers qualitatively appropriate results. On the other hand, companies are often afraid of leaving important decisions to an AI without (fully) understanding its decision-making process. The process model developed in the KI-AGIL project addresses these perceived risks through the explorative approach, in which companies are introduced in developing even complex AI solutions in small subsequent steps.

To reduce concerns about the complexity of AI development, the technical and organisational implementation of AI solutions and the necessary prerequisites were also discussed in workshops within the company networks.

2.4 Conclusion

At the start of the research project KI-AGIL, it quickly became apparent that companies with little experience in developing their own AI solutions had specific requirements regarding project planning and execution. Generally, these companies were able to outline potential application scenarios, but these were often very rough descriptions. In addition, the desired solutions were complex and demanding, and therefore not immediately solvable for the companies.

The analysis of existing agile models such as Scrum showed that they did not adequately consider the specifics of AI development. At the same time, models used specifically for AI development, such as CRISP-DM and CRISP-ML(Q), lacked a higher-level planning component that supports an iterative and explorative approach.

The process model developed in the KI-AGIL project attempts to close this gap. For this purpose, the project initially runs through a preceding phase in which possible fields of application are defined and compared with each other. As a result of the initial vision determination phase, a vision statement is formulated. This describes, at a relatively high level of abstraction, the medium-term goal to be achieved with AI. Based on this vision statement, concrete use cases are then defined in successive sprints to gradually approach the fulfilment of the vision statement. Important in this context are the feedback loops that take effect after each development phase and continuously during the productive use of the (partial) solutions. In these feedback loops, the results achieved and the knowledge gained during the course of the project are used to further refine the vision statement. It may also be necessary to reformulate the vision statement itself. Reasons for this could be changed external conditions or a different assessment due to an increased understanding of the process.

Within the scope of the research project KI-AGIL, the developed process model was successfully field-tested in two of the associated company projects. Here, the model proved that it allows a small-step approach to a complex AI solution and at the same time offers the possibility to sharpen the objectives after each development step or to adapt them if necessary.

Bibliography

Begleitforschung Mittelstand Digital (2021). Künstliche Intelligenz im Mittelstand. So wird KI für kleine und mittlere Unternehmen zum Game Changer. https://www.mittelstand-digital.de/MD/Redaktion/DE/Publikationen/ki-Studie-2021.pdf?__blob=publicationFile&v=1 (accessed 22 August 2023).

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T.P., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. https://www.researchgate.net/publication/225070403_CRISP-DM_10_step-by-step_data_mining_guide (accessed 22 August 2023).

Deloitte (2021). Künstliche Intelligenz im Mittelstand. https://www2.deloitte.com/content/dam/Deloitte/de/Documents/Mittelstand/Erfolgsfaktorenstudie_K%C3%BCnstliche%20Intelligenz%20im%20Mittelstand.pdf (accessed 22 August 2023).

Schröer, C., Kruse, F., Marx Gómez, J. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526–534. http://dx.doi.org/10.1016/j.procs.2021.01.199 (accessed 22 August 2023).

Schwaber, K., & Sutherland, J. (2020). The Scrum Guide. https://scrumguides.org/docs/scrumguide/v2020/2020-Scrum-Guide-US.pdf (accessed 22 August 2023).

Studer, S., Bui, T., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., Mueller, K. (2021). Towards CRISP-ML (Q): A machine learning process model with quality assurance methodology. https://doi.org/10.48550/arXiv.2003.05155 (accessed 22 August 2023).

Theis, S., Jentzsch, S., Deligiannaki, F., Berro, C., Raulf, A.P., Bruder, C. (2023). Requirements for explainability and acceptance of artificial intelligence in collaborative work. In: Degen, H., Ntoa, S. (eds) Artificial Intelligence in HCI. HCII 2023. Lecture Notes in Computer Science, vol 14050. Springer, Cham. https://doi.org/10.1007/978-3-031-35891-3_22 (accessed 22 August 2023).

Vogel, W. (2021). Agile Methods for Complexity management: A literature overview. In: Fritzsche, R., Winter, S., Lohmer, J. (eds), *Logistik in Wissenschaft und Praxis*. Springer Gabler, Wiesbaden. https://doi.org/10.1007/978-3-658-33480-2_6 (accessed 22 August 2023).

Robert Menger, Karla Ohler-Martins, Amanda Lemette,
Robert J. Martin

Chapter 3
Automatic Classification of Files Based on the Classes of IEC 61355

Abstract: The greatest value for plant operators is rapidly available, correctly classified, and designated documentation. For a typical power plant, for example, over 100,000 single files need to be classified manually by professionals taking approximately 5 minutes per file. The documents are often stored in different formats and lack any uniform information (such as title or layout). Therefore, operators have a great interest in automating these classification and designation processes. Within a joint research project, the Menger Engineering GmbH, the Ruhr West University of Applied Sciences, the Universität Duisburg-Essen and the Pontifical Catholic University of Rio de Janeiro cooperate to find solutions based on artificial intelligence (AI) in order to automatically classify and designate the files and extract specific information to support the maintenance processes. In this paper, we will introduce the current methods and results according to the classification of single files based on the classes of the IEC 61355, the most common standard for technical document classification.

Keywords: IEC 61355, IEC 81346, document classification code, file classification, technical document, DCC, reference designation system, document classification

3.1 Introduction

Nowadays, it is very important for operators of technical plants to have effective access to all the information and documents of the plant that are needed for operation and maintenance processes. Additionally, there are several legal requirements (depending on the kind and the location/country of the plant) which obligate the operators to provide their staff with all information about the plant required for a safe operation between human, machine and environment. Currently, engineers spend approximately 30% of their time searching for documents and information (Hahm,

Robert Menger, Menger Engineering GmbH, Leipzig, e-mail: robert@menger.group
Karla Ohler-Martins, Institute of Business Administration, Ruhr West University of Applied Sciences, Mülheim, e-mail: karla.ohler-martins@hs-ruhrwest.de
Amanda Lemette, Pontificia Universidade Catolica do Rio de Janeiro, Rio de Janeiro, e-mail: amanda.lemette@puc-rio.br
Robert J. Martin, Faculty of Engineering, University of Duisburg-Essen, e-mail: robert.martin@uni-due.de

https://doi.org/10.1515/9783110790320-003

Lee & Suh, 2015). For an effective access to the documents and information, structuring, classification and designation processes are needed. To this end, currently two main international standards provide the framework conditions: The IEC 61355 provides the basic methods for identification and classification of documents and information independent of the type of the plant, while the IEC 81346 provides generic rules and classes for the structuring and identification of systems, sub-systems and components of plants. According to the experiences of the Menger Engineering GmbH, it is well known that in most of the technical plants around the world, the provided information and documents are incomplete and most of the plant knowledge is person related. The main reason for this situation is that the structuring and designation processes of the information and documents are done manually with extensive effort.

For an effective access to the information and documents of a plant, it is also necessary to identify each document in a proper way, and to bring the document in relation to the plant structure (see Figure 3.1). Currently, neither of these processes can be reliably automated. In our joint research, we first investigate how different machine learning methods can improve on the state of the art with respect to the specific task of *technical document classification* into the aforementioned IEC categories. In the following, we present our preliminary results for the classification according to the IEC 61355.

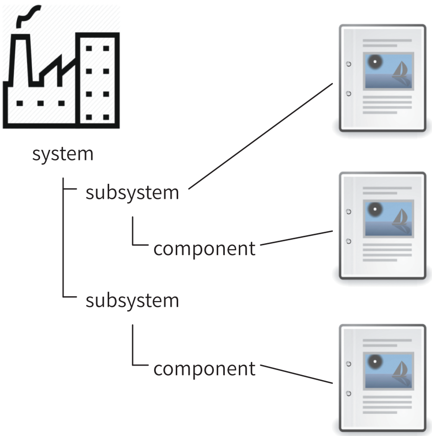


Figure 3.1: Designation of documents to the plant structure. (compiled by authors.)

3.2 The IEC 61355 Standard

The IEC 61355 – classification and designation of documents for plants, systems and equipment (International Electrotechnical Commission [IEC], 2009) – is an interdisciplinary, international standard, which contains rules and classification tables to identify a technical document independent from the manufacturer and the source systems.

There are three main arguments for the need of such a standard in the industry. First, these general basic rules for the identification of documents improve the communication and understanding between document interchanging parties, for example, the operator of a system and the several suppliers of the system parts. A second important aim of the standard is to provide the possibility to model the demand of documents based on the general rules and classes. A third reason for unifying the identification of documents is the correlation of documents and objects: The standard contains methods to designate the documents with an indication of the objects to which they belong.

Therefore, the IEC 61355 standard introduced a specific model of a document identifier (see Figure 3.2) which consists of three different parts:
- the content part (object designation),
- the class part (kind of document) and
- the counting part.

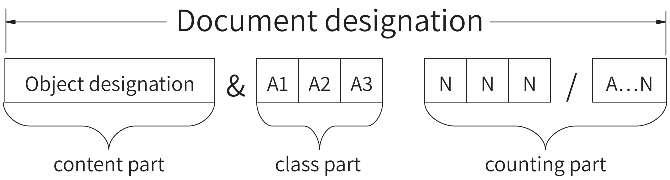


Figure 3.2: Document designation IEC 61355. (compiled by authors.)

Based on IEC 61355, a document can be described and identified by the object it contains information about, the class of the kind of the document and a number. In addition to the identification model, the standard contains specific classes for document kinds: the so-called Document kind Classification Code (DCC) contains 65 categories in total. A datasheet for a pump, for example, will be designated to the class "DA – Documents providing information on technical data and characteristics about material, products or systems that are necessary for their proper implementation" IEC (2009). The DCC contains a maximum of three letters, where the first letter is optional and represents the technical area followed by two letters for the document kind.

3.3 Related Work

The general task of automatically classifying documents has been the subject of extensive research in the past. In their seminal 1963 work, Borko & Bernick (1963) investigated the classification of documents in a mathematical way. Their factor analysis technique was able to classify a total of 90 computer literature abstracts in five categories with an accuracy of around 48%. This was the first demonstration of the feasibility of the classification of documents in an automatic way.

Researchers often investigated the automatic classification of *technical* documents based on key phrases depending on their frequency in text (Trappey et al., 2006). Good results were reached for the classification of technical documents of a patent database based on the combination of image and text analysis (Jiang et al., 2022). The visual and text analysis of documents is very similar to the human approach and gives models the possibility to define criteria for both the layout and the content. Similar techniques have been employed successfully for improving the classification accuracy by means of image-integrated text content analysis (Karagoglu et al., 2022).

Other recent studies investigated the concept of using transformer networks instead of convolutional networks for image classification tasks (Dosovitskiy et al., 2020) In Summer 2022 the Document Image Transformer (DiT) was introduced (Li et. al., 2022). This self-supervised pre-trained transformer model was pre-trained based on the IIT-CDIP Test Collection 1.0 (Lewis et al., 2006) with a large number of files classified according to a handful of criteria (contains tables or images, is handwritten, etc.). With the Bidirectional Encoder representation from Image Transformer (BEiT) (Bao et al., 2022) inspired representation of the images as image patches and visual tokens for pre-training, the DiT should be considered a good base model for the task of classifying technical documents.

Generally, prior work on technical document classification has focused on a large database of patents as a primary example dataset. In contrast to this project, all the classified documents of the patent database have a uniform title and abstract, which simplifies the classification task enormously.

3.4 Process of Classification

3.4.1 Methods and Available Data

The classification of documents according to the categories of the IEC standard IEC (2009) consists of two separate subproblems. As aforementioned, a single class code contains of three letters, where the first letter represents the *technical area* (see Figure 3.3) and the remaining two letters indicate the *document kind*.

We therefore separate the automatic classification of the files into two main tasks:

1. Classification of the technical area (one-letter-code).
2. Classification of the document kind (two-letter-code).

The technical area is theoretically represented by a total amount of six classes based on the standard IEC (2009). Within this project, we concentrate on the four classes

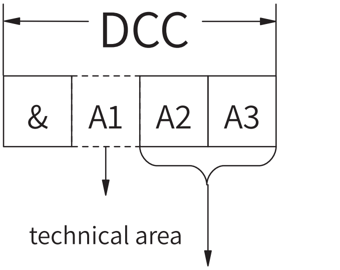


Figure 3.3: Structure of the DCC. (compiled by author.)

mainly used by the industry. Similarly, for the classes of the document kind, we will consider the most commonly used 39 categories. For both classification tasks, we will

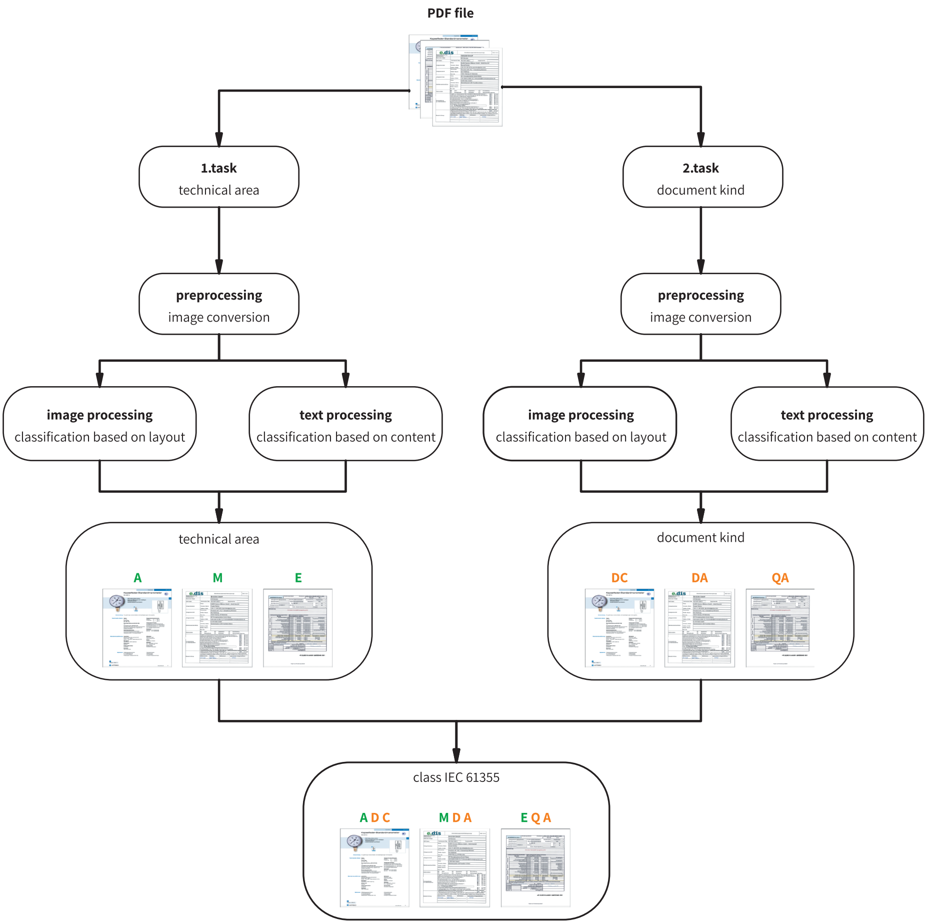


Figure 3.4: Main classification process. (compiled by authors.)

employ a combination of visual and content analysis. Apart from the text content, meta-information such as the number of pages can be used for the classification as well. In the preliminary work presented here, however, a purely image-based approach is applied in order to provide a first baseline.

At the beginning of the classification process (see Figure 3.4), we therefore need to implement a conversion of single- and multi-page PDF files to images. After the images are automatically classified based on the technical area and based on the document kind. Finally, the output of both classification processes can be combined to the real class of the IEC standard IEC (2009).

3.4.2 Training Procedure

The datasets for the training of the models are derived from projects of the Menger Engineering GmbH. This company is a service partner for the structuring processes of information for technical plants. During its approach to several projects in the past, related to different kinds of plants, the Menger Engineering GmbH has structured, classified and designated a very large number of documents and data, which can be provided for training and testing. In Table 3.1, the balance of total files in comparison to the different kind of plants is listed. The project step of automatically classifying the files according to the IEC 61355 standard, however, is independent of the kind of plant from which the file was obtained.

Table 3.1: Training dataset.

Kind of plant	Total of files
Renewable power plants	63,997
Waste fired power plants	51,390
Coal/gas fired power plants	12,625
Chemical plants	13,489
Gas storage plants	263,029
Total	**404,530**

The distribution of the files according to the technical area is mainly focused on civil, mechanical, and electrical engineering (see Figure 3.5). This scenario is due to the actual use of the classes in the industry.

In the term of document classes there is also a concentration of just a few classes according to our total amount of files (see Figure 3.6).

The different types of files (CAD files, MS office files, image files, PDF files) need to be transferred to a unified PDF format before using them for the classification process. For the first experiment, a smaller dataset of 10,000 files per class (except class

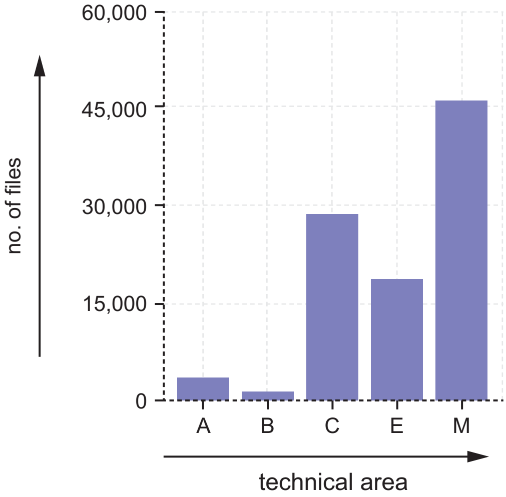


Figure 3.5: Distribution of the technical area. (compiled by authors.)

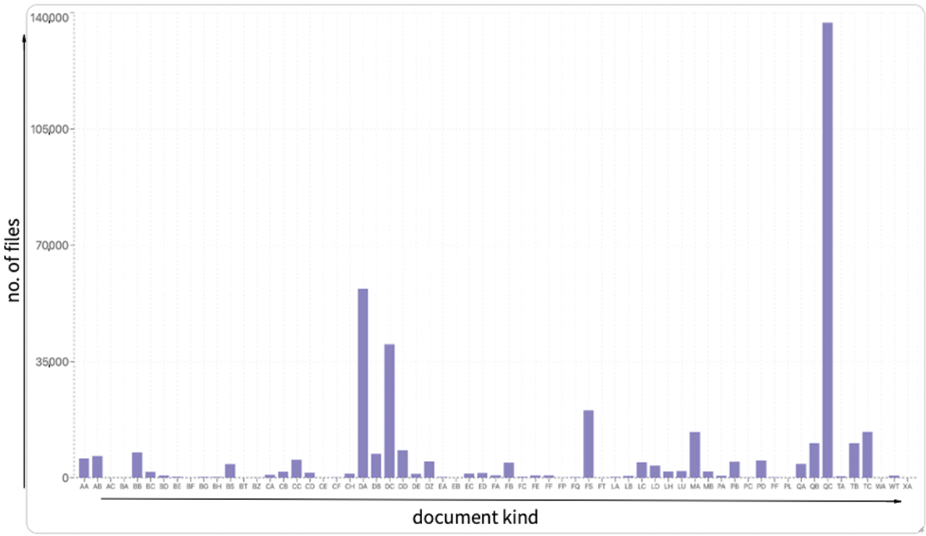


Figure 3.6: Distribution of document kinds. (compiled by authors.)

"A" with just 2,395 files) for the first task (technical area) and a perfectly balanced set of 500 files per class for the second task (document kind) was selected.

3.4.3 Preliminary Experiments

For the first experiment, the small dataset for the first classification task (technical area) was converted to images in three different ways:
1. Single-page conversion
2. Multi-page conversion
3. Cropped-part conversion

Here, single-page conversion means that the image was based on only one page per PDF, independent of the number of pages of the file. To avoid inconsistency because of cover sheets and tables of contents, in the case of a multi-page PDF file, the third or the fifth page was converted to an image.

To preserve more information from each document in the dataset, we also used alternative methods of converting more than one page in case of multi-page files. First, for the multi-page conversion, a maximum of five separate images per multi-page PDF file was generated. The limitation of five was chosen for not unbalancing the dataset too much.

Finally, another set of data was generated by randomly cropping quadratic fixed-sized parts from multi-page files and combining them into a single image file. As a result, we obtain one image file representing the content of a multi-page PDF file. For the training process, all the images were scaled to a maximum of 512 px for the longest side.

The datasets were divided into training (70%), validation (20%) and test data (10%) before the conversion process to avoid overlapping of the PDF files between the training and the test data.

These three datasets were then used to train both a classical convolutional network (CNN) and a pre-trained transformer model. The CNN consists of an input layer with a normalisation function, a total of three convolutional layers (16, 32, 64 neurons) with a kernel size of three and a rectified linear unit activation function, and two dense layers (128, 4 neurons).

For the transformer model, we used the Document Image Transformer (Li et al., 2022), a self-supervised transformer model pre-trained on 42 million images of different documents. For our first studies, we use the default parameters of this model.

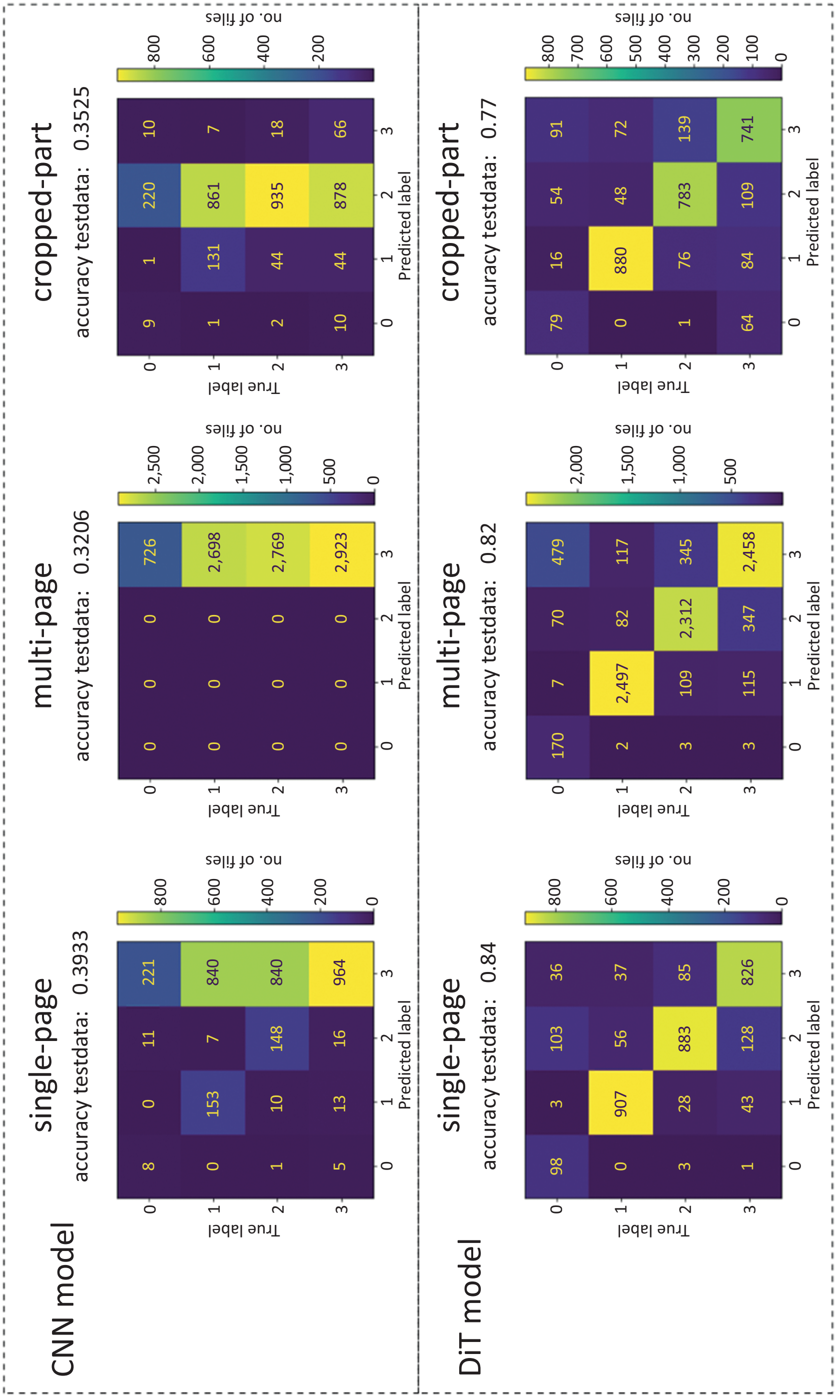


Figure 3.7: Results of the training and test phase of CNN in terms of accuracy and confusion matrices. (compiled by authors.)

3.4.4 Results

To evaluate the suitability of a purely image-based categorisation, we first used the three prepared datasets described in the previous section to train the classical CNN. The only hyperparameter optimisation consisted of changes to the learning rate and the total number of training epochs. The optimum was found with between 6 and 10 epochs and the learning rate of 0.0001. However, as shown in Figure 3.7, the results of the three different training processes with the single-page, the multi-page and the cropped-part conversion demonstrate that the network was unable to reliably identify the correct class labels. The accuracy for the trained models related to the test data was between 30% and 40%. While the single-page dataset seems to perform slightly better than the alternatives, it should be noted that the prediction is almost constant in all three cases, with only a minority of the other classes being categorised correctly even in the single-page case (A). With the application of the transformer model (DiT; Li et al., 2022) based on the default properties, the same prepared datasets obtain accuracies between 80% and 84%. In addition, the confusion matrices show a much better performance of the transformer network compared to the generic convolutional one. It should be kept in mind that the total number of files was much lower for the first class than for the other classes, which can explain the lower accuracy for this technical area label, as shown by the top rows of the confusion matrices. The worst results are achieved with the cropped-part images, while the best case for the classification was the single-page dataset.

3.5 Conclusion and Future Work

Our experiments show that a classical generic CNN model applied to the documents as pure images is not immediately suitable for the classification process without any further optimisation: for all datasets, the model was focusing on one class out of the four, which resulted in a low accuracy of around 40%. In contrast, the pre-trained transformer network (Li et al., 2022), on the same base of datasets, obtained an accuracy of over 80%, which is a good base for future experiments. In addition, it was shown that the best results were based on the single-page converted images; expanding the dataset by generating more than one image from each PDF file did not produce any improvements according to the training results.

In future experiments, we will focus on optimising the results of the classification process by expanding the input space of the neural network to include additional properties of the PDF files as well as the textual content of the files. In addition, our methods will be applied to the classes for the document kind for the second classification task.

Bibliography

Bao, H., Dong, L., Piao, S., & Wei, F. (2022). BEiT: BERT Pre-Training of Image Transformers. http://arxiv.org/pdf/2106.08254v2.

Borko, H., & Bernick, M. (1963). Automatic document classification. *Journal of the ACM*, 10(2), 151–162. https://doi.org/10.1145/321160.321165.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., & Unterthiner, T. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. http://arxiv.org/pdf/2010.11929v2.

Hahm, G. J., Lee, J. H., & Suh, H. W. (2015). Semantic relation based personalized ranking approach for engineering document retrieval. *Advanced Engineering Informatics*, 29(3), 366–379. http://dx.doi.org/10.1016/j.aei.2015.01.003.

International Electrotechnical Commission. (2009). Classification and designation of documents for plants, systems and equipment. International Standard IEC 61355-1:2009.

Jiang, S., Hu, J., Magee, C. L., & Luo, J. (2022). Deep learning for technical document classification. *IEEE Transactions on Engineering Management*, 1–17. https://doi.org/10.1109/TEM.2022.3152216.

Karaoglu, S., Tao, R., Gevers, T., & Smeulders, A. W. M. (2017). Words matter: Scene text for image classification and retrieval. *IEEE Transactions on Multimedia*, 19(5), 1063–1076. https://doi.org/10.1109/TMM.2016.2638622.

Lewis, D., Agam, G., Argamon, S., Frieder, O., Grossman, D. & Heard, J. (2006). Building a Test Collection for Complex Document Information Processing. Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. https://doi.org/10.1145/1148170.1148307.

Li, J., Xu, Y., Lv, T., Cui, L., Zhang, C., & Wei, F. (2022). DiT: Self-supervised Pre-training for Document Image Transformer. DOI: http://arxiv.org/pdf/2203.02378v3.

Trappey, A. J. C., Hsu, F., Trappey, C. V., & Lin, C. (2006). Development of a patent document classification and search platform using a back-propagation network. *Expert Systems with Applications*, 31(4), 755–765. https://doi.org/10.1016/j.eswa.2006.01.013.

Part 3: **Finance and Accounting**

Marcus Bravidor

Chapter 4
Auditing Algorithms in the (Non-)Financial Audit: Status Quo and Way Forward

Abstract: In this chapter, we analyse how algorithmic audits could be integrated into (non-)financial audits. We compare auditing approaches for learning (dynamic) and rules-based (static) algorithms. It is questionable whether learning algorithms can even be audited as stand-alone or standard software. Our results indicate a clear gap in the scope, risk assessment and audit procedures applicable to learning algorithms. Whereas the general rules for auditing static algorithms also apply to learning algorithms, these cannot account for the particularities of the latter. We highlight different frameworks for the auditing of learning algorithms and provide an outlook on open challenges for standard-setters and users.

Keywords: artificial intelligence, auditing, auditing algorithms, auditing machine learning

4.1 Introduction

In the last couple of years, a lot of debate in the (non-)financial audit research and practice community has focused on the application of big data and machine learning (ML) tools in the audit process (Cao, Chychyla & Stewart, 2015). Automation and data analytics are expected to improve audit quality and efficiency. Therefore, audit work will rely less on sampling procedures and more on automated evaluations of the underlying business processes. Whereas business processes per se are technology-neutral, they are often supported or partially performed by information systems (e.g., enterprise resource planning systems). As far as these processes build upon retrograde information, such as financial transactions, automatic processing is straightforward. However, many tasks in the (non-)financial reporting domain are less well-structured and require judgement, hence, introducing discretion. Examples are non-repetitive transactions (e.g., provisions) and any form of forward-looking information, which is particularly relevant to stakeholders (Francis, Hanna & Vincent, 1996; Siems, Seuring & Schilling, 2023). Against this background, it is surprising how little attention has been paid to incorporating algorithm audits into (non-)financial audits.

Marcus Bravidor, University of Freiburg, Financial Accounting & Auditing, Freiburg, Germany, e-mail: marcus.bravidor@accounting.uni-freiburg.de

https://doi.org/10.1515/9783110790320-004

In this chapter, we analyse how algorithmic audits could be integrated into (non-) financial audits. In doing so, we combine two strands of research. First, we outline the status quo of (accounting) information systems audits. Second, we build upon the growing literature on audit procedures and frameworks for (self-learning) algorithms to highlight current shortfalls of and potential benefits for the (non-)financial audit (Bandy, 2021; Costanza-Chock, Raji & Buolamwini, 2022; Metaxa et al., 2021; Vecchione, Levy & Barocas, 2021). Combining both strands, we compare auditing approaches for learning (dynamic) and rules-based (static) algorithms. It is questionable whether learning algorithms can even be audited as stand-alone or standard software. This approach seems reasonable since (learning) algorithms do not only gain importance for (non-)financial reporting but for management and business models in general. From a reporting point of view, potential risks associated with biased or erroneous algorithms have first-order effects for numbers represented in financial reports and second-order effects for other material information (e.g., risk reporting).

This chapter contributes to several strands of literature. First, the governance of algorithms will be in the spotlight with the European Union's upcoming Artificial Intelligence Act (European Commission, 2021; Mökander et al., 2022). Questions of which algorithms will have to be audited, by whom and how will gain even more practical importance. Second, the intersection of (non-)financial auditing and algorithm auditing has yet to be addressed by the accounting information systems literature as well as the respective standard-setters (Cho et al., 2020). To our knowledge, this chapter is the first to analyse if such an integration is feasible and how it could be done.

The chapter is structured as follows. Section 2 provides a short introduction to the financial audits based on the International Standards on Auditing (ISA). The nature of algorithms, in particular the difference between static and learning algorithms is covered in Section 4.3. Audit approaches from auditing standards and literature to address different types of algorithms are introduced in Section 4.4. Section 4.5 lists some open issues for the integration of learning algorithm audits in financial audits. A quick look forward forms the conclusion in Section 4.6.

4.2 (Non-)Financial Audits

Financial audits provide external validation for the financial information prepared by firms. This validation is expressed in the auditor's concluding opinion on whether the financial information provides a true and fair view of the firm's position and is free from material errors. The International Standards on Auditing (ISA) are a broad set of standards that covers most areas of the process necessary to reach such a conclusion. The process encompasses several steps: client acceptance, audit planning, assessing the risk of material misstatement, developing a risk response, performing the actual risk response and audit procedures, including the gathering of evidence as well as forming and reporting an opinion based on these findings (Arens et al., 2021).

The risk assessment is based on two types of risk (ISA 200.13(n)). Inherent risk is the susceptibility of any assertion about a transaction, balance or disclosure to a material misstatement. Control risk is the risk that such a material misstatement would not be detected or corrected by the internal control systems. Auditors need to test these internal controls with reasonable audit procedures. The design of these procedures lies with the discretion of the auditor. ISAs are, per se, technology neutral, i.e., they do not promote any specific audit technology or provide guidelines for certain technical implementations of controls (Gantz, 2014; Gilewicz, 2022). However, evidence for the effectiveness of internal controls can also be gathered in upstream audits, e.g., internal or software audits (ISA 330.13(a); ISA 500.A18; ISA 500.A51). In summary, international auditing standards do not provide any specific guidance on the handling of algorithms in general but indirectly acknowledge them as far as those concerning the financial reporting process.

4.3 The Nature of Algorithms

Against popular belief, algorithms are not necessarily related to technology: "Informally speaking, an algorithm is a collection of simple instructions for carrying out some task. Commonplace in everyday life, algorithms sometimes are called procedures or recipes." (Sipser, 2013, p. 182). To fulfil this task, algorithms need some clearly defined properties (Schneider & Gersting, 2010): well-ordered (correct order to execute the single operations), unambiguous (operations are clear and cannot be simplified), effectively computable operations (every operation must be possible), produce a result (that can be judged as correct or not), and halt in a finite number of time (have some operation that stops the algorithm).

Based on this broad definition, algorithms can be found in every software, data analytics and ML program. Theoretically, every well-designed process shows some resemblance to an algorithm. One example is the definition of business process by Davenport and Short (1990, p. 13) as "a set of logically-related tasks performed to achieve a defined business outcome". In this regard, business processes are "deterministic machines" (Melao & Pidd, 2000, p. 112) that will always produce the same output given the same input. Additionally, they are static in the way that changes in their behaviour are exogenous to them. Any changes that affect the single operations, their order or connection cannot be a result of the algorithm itself but of an external force (e.g., programmer). Such tasks that consist of logically linked and sequential operations are common in accounting information systems (e.g., purchase to pay- or order to cash-cycles).

ML, as a subfield of artificial intelligence, also fits the definition of an algorithm. However, machine-learning models deviate from the aforementioned since they “learn” from data to leverage their performance (Mitchell, 1997). In that sense, they are dynamic. Inputs are not only processed but can directly affect the behaviour of the model. For supervised models, input data is separated into training and test data. The models learn based on the test data and apply the learned patterns to the test data. For unsupervised models, this separation is not necessary. The model is “blind” and adjusts based on input data alone. However, it is important to note that the performance of dynamic algorithms does not only rely on the technical proficiency of the algorithm alone but also on the quality of the training and/or input data. Letting algorithms learn may increase the performance but comes at the cost of potential downsides like biases (Kordzadeh & Ghasemaghaei, 2022). Biases can be roughly defined as any distinct source of any unwanted outcome in an ML model that has the potential to cause harm (Hellström, Dignum & Bensch, 2020; Suresh & Guttag, 2021). These problems exacerbate if models are used for advanced problem-solving, often termed as artificial intelligence, such as generating predictions, rules, answers or recommendations (Russell & Norvig, 2021).

4.4 Auditing Algorithms

4.4.1 Static Algorithms

As seen before, business processes, which are static algorithms, are the backbone of computer-based accounting information systems. From an auditing perspective, the procurement or development, and deployment of such information systems is subject to the firms’ compliance management, and internal control organization. Relevant frameworks (e.g., COSO) distinguish general and application controls (Hall, 2011). Application controls ensure the validity, completeness and accuracy of financial transactions. General controls encompass all activities that do not relate to specific transactions but the systems as a whole.

A well-defined and documented systems development life cycle is one of the most important general controls. It ensures that new systems are aligned with organizational strategy and the related tasks, work efficiently and free from material errors and, hence, improve the control environment (Hall, 2011; Isaias & Issa, 2015; ISO, 2015, 2017). There are several iterations of system (and software) development life cycles. The actual design depends on the organization. The usual phases are planning, design, programming and testing, implementation and monitoring. Audits may pick up on any of these stages to test whether relevant life cycle stages have been followed as well as to test or review whether relevant application controls have been implemented (Gantz, 2014).

On an application level, auditors need to consider the typical kinds of error and fraud—and adequate control procedures (e.g., data editing, documentation and manuals, staff handling of data input and output, and effective error correction procedures). Audit procedures encompass system review and a test of controls. Whereas reviews are usually based on the documentation and help assess potential risks, evidence on the effectiveness of controls is gathered through program evaluation. This means that test data with known results is fed into the algorithm. Other approaches are audit hooks (e.g., identifying questionable transactions) or review of system snapshots or log files (Romney & Steinbart, 2018).

4.4.2 Dynamic Algorithms

Dynamic algorithms possess unique challenges to audits. Le Merrer, Pons, and Trédan (2022) highlight this with a metaphor from police work and criminology. Audits of static algorithms follow well-defined characteristics of desirable and undesirable behaviour of an algorithm. Just like a policeman who has precise regulations on whether a car is parked correctly or not. This "bobby audit form" checks inputs against outputs but requires a precisely formulated list of potential behaviours. The more heterogenous these behaviours are, the more likely it becomes that certain aspects are not covered or cannot be tested due to budget constraints. They propose an alternative, the "Sherlock audit form" which does not only separate right from wrong but builds an entire case or narrative (e.g., including motives, absence of alibis). In other words, the audit can go as far as a test if forged data could train the algorithm to present an undesired result. However, even this approach does not provide guidance on the frequency and the single stages of the algorithm life cycle an audit should deal with.

One stream of research adapts life cycle models from software and systems development to address this issue. These models are—like financial audits—based on a risk-based approach.

Raji et al. (2020) propose the SMACTR framework for internal algorithm audits. This life cycle framework is based on a risk assessment that captures the social and ethical effects of the algorithm in six stages: scoping, mapping, artefact collection, testing, reflection and post-audit. Boer, Beer & van Praat (2023) suggest a threefold assessment of the risk of a particular algorithm: complexity (how hard it is to predict the impact of the algorithm on a system), autonomy (degree of human interference in the execution of the algorithm) and impact (direct effect of the algorithm on the financial or information position, attention, rights, duties or powers of individuals, groups or the organization). In essence, the assessment should capture the economic, social and ethical effects of the algorithm. The first stage, scoping, builds upon these results and aims to investigate the areas and potential sources of harm through a description of the system use case and an impact assessment. In the second stage, mapping, key stakeholders and collaborators who are affected by the audit and/or are necessary for the execution of the audit should be identified and mapped. Where as this stage focusses on the organizational side, the third stage, artefact collection, concerns the collection of first evidence by gathering the documentation of the development process. Further audit procedures, which are collected in a checklist, are based on opportunities for testing identified and prioritized in this stage. The actual testing is done in the following stage and is based on scoping and the checklist. Testing activities are executed to ensure compliance with the (ethical) values that the system should comprise. This results in an Ethical Risk Analysis along two dimensions: (1) likelihood and (2) severity of a risk/failure. In the final reflection stage, the risk assessment is updated based on the testing results to include an outline of specific organizational principles that the algorithm could affect. Auditors prepare a summary report as well as a remediation and a risk mitigation plan. However, Brown, Davidovic & Hasan (2021) point out that there is always a trade-off between different metrics and stakeholder interests. So, any results may affect not only the risk assessment but also the assumptions it is built upon.

Life cycle models provide a more comprehensive overview of the single stages of an algorithm audit. However, the audit life cycle works in parallel to the software development life cycle. Hence, both cycles complement each other. The SMACTR model, for example, is retrospective in the sense that a working version of the algorithm is subject to the audit. Given the fast-changing nature of learning algorithms, Mökander and Axente (2022) propose the definition of intervention points for (ethical) deliberations. Their approach is more grounded in a constant control framework that addresses the entire organization and must be derived from known and unknown shortcomings of the corporate culture (e.g., incentive structures, trade-offs in the design phase, communication channels to continuously monitor and redesign algorithms). Put differently, an effective audit is incorporated into the software development lifecycle. Sandu, Wiersma & Manichand (2022) combine these approaches by introducing a lifecycle framework for audits that follows model risk management based on the Three Lines of Defense (The Institute of Internal Auditors, 2020). The algorithm audit is defined as a continuously improving cycle with two phases: before use (initiation, development, implementation) and during use (use, monitoring, review, retirement). Within this cycle, a constant validation process provides updates on the inclusion and importance of already identified and new risks.

Life cycle models are helpful in identifying the material steps as well as supporting risk assessment and mitigation. They, however, fall short on providing tools and procedures to conduct the actual audit. Koshiyama, Kazim & Treleaven (2022) provide a two-way approach to fill this gap. In their framework, the landscape of algorithm auditing depends on the level of access to the algorithm and the target-state dimensions. Access to the algorithm depends on the auditors' situation within the organization and the risk of an algorithm. Voluntary audits, in particular, may not have full access to data and code ("black box") or have resource restrictions that prohibit full-scale access ("white box"). Hence, auditors may either choose a certain level of access (process access aka black box, model access, input access, outcome access, parameter control, learning goal, white box) or be presented with one. On the other hand, target-state dimensions do not directly depend on the level of access, but the audit approach is contingent on the level of access. The importance of target state dimensions within the audit depends on the risk assessment. Target states are explainability, robustness, fairness, privacy, information concealed, feedback detail, typical applications and appropriate oversight. The audit results are formed into mitigation strategies to comply with any standards the algorithm is subject to.

Overall, there are a considerable number of frameworks on algorithm audit planning, risk assessment and potential audit procedures. Currently, these approaches focus on an intra-organizational perspective and underline the importance of effective internal control systems over the life cycle of an audit.

4.5 Integration of Algorithm Audits in the (Non-)Financial Audit

The benefit of technology-neutral audit standards is their flexibility towards innovation and new technological developments. The drawback is the lack of clear guidance when such developments occur—as is the case with learning algorithms. As outlined above, the audit of learning algorithms can be incorporated into the general internal control framework that auditors rely upon. Thus far, the approach is quite like the one for static software. Furthermore, most of the aspects of learning algorithms are already part of other audit procedures (e.g., data, security, privacy, deployment). Against this background, it may be questionable whether explicit rules for financial auditing are even necessary. Internal audit is timelier and has easier access to different stages of the life cycle. On the other hand, external auditors provide credibility due to their independence and are also a fall-back option if firms lack the internal resources or competencies. External auditors, hence, need standards to be a viable fall-back option and to judge whether internally carried out audit procedures have been sufficient. To develop such a full-fledged audit standard, some open areas remain:

Scope: The outlined approach for learning algorithms has a strong focus on their ethical and social implications (Bandy, 2021; Brown, Davidovic & Hasan, 2021; Koshiyama, Kazim & Treleaven, 2022; Metaxa et al., 2021). Many of these aspects are important for society or the protection of individuals (e.g., credit scoring) and society. Financial audits, on the other hand, primarily focus on the true and fair view of financial information and the related business processes. In that case, other implications are of minor concern if they do not affect the (financial) risk position of the organization. In light of the increasing importance of environmental, social, and governance (ESG) criteria and the call for corporate digital responsibility (Lobschat et al., 2021; Mueller, 2022), a broader focus may be imminent.

Risk assessment: Audits of learning algorithms can have different directions (as outlined by target-state dimensions in Kazim & Koshiyama, 2021) that are not mutually exclusive (e.g., fairness and explainability). The risk assessment, on the other hand, depends on the critical aims and scope. It is most likely that algorithm audits will not have a single aim (e.g., financial reporting) but a "multi-focused success criterion" (LaBrie & Steinke, 2019).

Audit procedures: As stated before, for many aspects certain audit routines and methods are already in place. Additionally, auditors can rely on other assurance services such as internal or software audits (Galli & Calzolari, 2021). To decide whether these alternative sources of assurance provide reasonable evidence, formal guidelines on risk assessment and related procedures are necessary. The relevant audit procedures are contingent on the risks related to the success criteria (LaBrie & Steinke, 2019). The framework from Koshiyama, Kazim & Treleaven (2022) could provide a starting point for an in-depth discussion since risk assessment and different levels of access require different levels of assurance. However, depending on the type and scope of the audit, auditors may lack access to data or face legal boundaries (e.g., data scraping). Disclosures and mandatory peer review rules may aid these shortcomings (Costanza-Chock, Raji & Buolamwini, 2022).

4.6 Concluding Remarks

So far, specific official and legally binding guidance for the consideration of learning algorithms in the (non-)financial audit is missing. Put differently: "Standards are thin, at best" (Costanza-Chock, Raji & Buolamwini, 2022, p. 1576). However, the standard rules for static algorithms apply in these cases as well. To close the remaining gap, auditors rely on custom frameworks and tools. This leads to the first open question: Are new standards necessary in the first place? As shown above, there is a broad set of standards, and some standard setters have already put up guidance on how learning algorithms affect their standards. For example, high-level guidelines for COBIT were published back in 2018 (ISACA, 2018). Given the broad scope of issues that ranges from the aim of the audit to the technicalities, additional guidance would be advisable but is not in sight. The current proposal for the 2024–2027 work plan of the International Auditing and Assurance Standards Board (IAASB) makes no reference to artificial intelligence; its predecessor (2020–2023) did but focused on the use of technology in the audit, not the audit of technology per se (IAASB, 2020, 2023). Even though algorithms are not a direct subject of ISA, auditors are subject to them in a multitude of ways. They are users (e.g., with computer-assisted audit technologies) and first and foremost, they need to ensure that such systems work reliably, at least concerning financial transactions. Pushes for clearer guidance (ISACA, 2018; Mökander & Axente, 2022) are therefore understandable. Against this background, future research could look at the cross-section of algorithm audits with other contemporary developments; i.e., agile auditing or continuous auditing which may be better suited to deal with this novel kind of fast-changing and always-developing algorithms. For so long, auditors do not need to worry. They dealt with changes like e-commerce or cloud computing before. And, if anything, this may be proof that humans remain an important aspect of the audit task (Alexiou, 2021). They have the unique ability to understand and assess the business risk, critically interpret the results of algorithms and identify implausibilities.

Bibliography

Alexiou, S. (2021). Algorithms and the auditor. *ISACA Journal*, 6, 1–8. https://www.isaca.org/resources/isaca-journal/issues/2021/volume-6/algorithms-and-the-auditor.

Arens, A. A., Elder, R. J., Beasley, M. S., Hogan, C. E., Jones, J. C., & Amerski, B. M. (2021). *Auditing: The Art and Science of Assurance Engagements* (Fifteenth Canadian edition). Pearson.

Bandy, J. (2021). Problematic machine behavior: A systematic literature review of algorithm audits. *Proceedings of the ACM on Human–Computer Interaction*, *5*(CSCW1), 1–34. https://doi.org/10.1145/3449148.

Boer, A., Beer, L. de, & van Praat, F. (2023). Algorithm assurance: Auditing applications of artificial intelligence. In: E. Berghout, R. Fijneman, L. Hendriks, M. de Boer, & B.-J. Butijn (Eds.), *Progress in IS. Advanced Digital Auditing: Theory and Practice of Auditing Complex Information Systems and Technologies* (1st edition, pp. 149–183). Springer International Publishing. https://doi.org/10.1007/978-3-031-11089-4_7.

Brown, S., Davidovic, J., & Hasan, A. (2021). The algorithm audit: Scoring the algorithms that score us. *Big Data & Society*, *8*(1), 2053951720983865. https://doi.org/10.1177/2053951720983865.

Cao, M., Chychyla, R. & Stewart, T. (2015). Big data analytics in financial statement audits. *Accounting Horizons*, *29*(2), 423–429. https://doi.org/10.2308/acch-51068.

Cho, S., Vasarhelyi, M. A., Sun, T., & Zhang, C. (2020). Learning from machine learning in accounting and assurance. *Journal of Emerging Technologies in Accounting*, *17*(1), 1–10. https://doi.org/10.2308/jeta-10718.

Costanza-Chock, S., Raji, I. D., & Buolamwini, J. (2022). Who audits the auditors? Recommendations from a field scan of the algorithmic auditing ecosystem. In: *2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1571–1583). ACM. https://doi.org/10.1145/3531146.3533213.

Davenport, T. H., & Short, J. E. (1990). The new industrial engineering: Information technology and business process redesign. *Sloan Management Review*, *31*, 11–27.

European Commission. (2021). Proposal for a Regulation of the European Parliament and of the Council laying down harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts, COM(2021) 206 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

Francis, J., Hanna, J. D., & Vincent, L. (1996). Causes and effects of discretionary asset write-offs. *Journal of Accounting Research*, *34*, 117. https://doi.org/10.2307/2491429.

Galli, N., & Calzolari, G. (2021). Auditing of Algorithms: Closed-Door Panel of the Florence Competition Summer Conference. https://ssrn.com/abstract=3972274.

Gantz, S. D. (2014). *The Basics of IT Audit: Purposes, Processes, and Practical Information*. Syngress, an imprint of Elsevier.

Gilewicz, G. (2022, April 14). *Technology's impact on the audit*. CPA Canada. https://www.cpacanada.ca/en/business-and-accounting-resources/audit-and-assurance/blog/2022/april/technologys-impact-on-audit.

Hall, J. A. (2011). *Information Technology Auditing* (3rd edition). South-Western/Cengage Learning.

Hellström, T., Dignum, V., & Bensch, S. (2020, April 1). *Bias in machine learning – What is it good for?* https://arxiv.org/pdf/2004.00686.

IAASB. (2020). *Strategy for 2020–2023*. New York, NY. https://www.ifac.org/system/files/publications/files/IAASB-Strategy-for-2020-2023-V6.pdf.

IAASB. (2023). *The IAASB's Proposed Strategy and Work Plan for 2024–2027*. New York, NY. https://www.ifac.org/system/files/publications/files/IAASB-Consultation-Strategy-2024-2027.pdf.

The Institute of Internal Auditors. (2020). The IIA's Three Lines Model: An update of the Three Lines of Defense (Lake Mary, FL). https://www.iia.org.au/sf_docs/default-source/technical-resources/the-iias-three-lines-model–an-update-of-the-three-lines-of-defence.pdf?sfvrsn=4.

ISACA. (2018). *Auditing Artificial Intelligence*. Schaumburg, IL. https://ec.europa.eu/futurium/en/system/files/ged/auditing-artificial-intelligence.pdf.

Isaias, P., & Issa, T. (2015). Information system development life cycle models. In: P. Isaias (Ed.), *High Level Models and Methodologies for Information Systems* (pp. 21–40). New York: Springer. https://doi.org/10.1007/978-1-4614-9254-2_2.

ISO. (2015). ISO 15288: Systems and software engineering – System life cycle processes. https://www.iso.org/standard/63711.html.

ISO. (2017). ISO 12207: Systems and software engineering – Software life cycle processes. https://www.iso.org/standard/63712.html.

Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns*, *2*(9), 100314. https://doi.org/10.1016/j.patter.2021.100314.

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, *31*(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212.

Koshiyama, A., Kazim, E., & Treleaven, P. (2022). Algorithm auditing: Managing the legal, ethical, and technological risks of artificial intelligence, machine learning, and associated algorithms. *Computer*, *55*(4), 40–50. https://doi.org/10.1109/MC.2021.3067225.

LaBrie, R. C., & Steinke, G. (2019). *Towards a Framework for Ethical Audits of AI Algorithms* (Twenty-fifth Americas Conference on Information Systems). Cancun. https://aisel.aisnet.org/amcis2019/data_science_analytics_for_decision_support/data_science_analytics_for_decision_support/24/.

Le Merrer, E., Pons, R., & Trédan, G. (2022, February 15). *Algorithmic audits of algorithms, and the law*. https://arxiv.org/abs/2203.03711.

Lobschat, L., Mueller, B., Eggers, F., Brandimarte, L., Diefenbach, S., Kroschke, M., & Wirtz, J. (2021). Corporate digital responsibility. *Journal of Business Research*. 122(January), 875–888. https://doi.org/10.1016/j.jbusres.2019.10.006.

Melao, N., & Pidd, M. (2000). A conceptual framework for understanding business processes and business process modelling. *Information Systems Journal*, *10*(2), 105–129. https://doi.org/10.1046/j.1365-2575.2000.00075.x.

Metaxa, D., Park, J. S., Robertson, R. E., Karahalios, K., Wilson, C., Hancock, J., & Sandvig, C. (2021). Auditing algorithms: Understanding algorithmic systems from the outside. *Foundations and Trends® in Human–Computer Interaction*, *14*(4), 272–344. https://doi.org/10.1561/1100000083.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill series in computer science. McGraw-Hill.

Mökander, J., & Axente, M. (2023). Ethics-based auditing of automated decision-making systems: Intervention points and policy implications. *AI & Society*. 38, 153–171. https://doi.org/10.1007/s00146-021-01286-x.

Mökander, J., Axente, M., Casolari, F., & Floridi, L. (2022). Conformity assessments and post-market monitoring: A guide to the role of auditing in the proposed European AI regulation. *Minds and Machines*, *32*(2), 241–268. https://doi.org/10.1007/s11023-021-09577-4.

Mueller, B. (2022). Corporate digital responsibility. *Business & Information Systems Engineering*. 64, 689–700. https://doi.org/10.1007/s12599-022-00760-0.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020, January 3). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. https://arxiv.org/abs/2001.00973.

Romney, M. B. & Steinbart, P. J. (2018). *Accounting Information Systems* (14th edition). Pearson.

Russell, S. J., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th edition). Pearson series in artificial intelligence. Pearson.

Sandu, I., Wiersma, M., & Manichand, D. (2022). Time to audit your AI algorithms. *Maandblad Voor Accountancy En Bedrijfseconomie*, *96*(7/8), 253–265. https://doi.org/10.5117/mab.96.90108.

Schneider, G. M., & Gersting, J. L. (2010). *Invitation to Computer Science* (5th edition). Course Technology Cengage Learning.

Siems, E., Seuring, S., & Schilling, L. (2023). Stakeholder roles in sustainable supply chain management: A literature review. *Journal of Business Economics*. 93, 747–775. https://doi.org/10.1007/s11573-022-01117-5.

Sipser, M. (2013). *Introduction to the Theory of Computation* (3rd edition, international edition). Cengage Learning.

Suresh, H., & Guttag, J. (2021, January 28). *A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle*. https://arxiv.org/pdf/1901.10002.

Vecchione, B., Levy, K., & Barocas, S. (2021). Algorithmic auditing and social justice: Lessons from the history of audit studies. In: *Equity and Access in Algorithms, Mechanisms, and Optimization*. ACM. https://doi.org/10.1145/3465416.3483294.

Isabel Lausberg, Arne Eimuth, Anne Stockem Novo

Chapter 5
Barriers to the Use of Artificial Intelligence (AI) in Management Reporting

Abstract: Management reporting as part of corporate management accounting serves to provide decision-makers with systematic, transparent, structured and up-to-date information (Weber, Malz & Lührmann, 2012). In corporate practice, management reporting is often experienced as a resource-intensive and time-consuming process, lagging behind in the adoption of technological innovations. The article explores how management reporting can be transformed into a state-of-the-art reporting that harnesses the potential of artificial intelligence (AI). Based on a comprehensive literature analysis and in-depth interviews with experts from academia and corporate practice, the authors propose a maturity model outlining a path to an AI-enabled management reporting. Building upon this model in combination with the Technology-Organization-Environment (TOE) framework, they further analyse the inhibitors and enablers of the dissemination of AI in management reporting.

Keywords: management reporting, maturity model, TOE framework

5.1 Introduction

Management reporting is a central part of management accounting and the finance function of companies. It serves to provide corporate management with systematic, transparent, structured and up-to-date information as a comprehensive basis for decision-making (Weber, Malz & Lührmann, 2012). In corporate practice, management reporting is often experienced as a resource-intensive and time-consuming process (e.g., Weißenberger, 2021). The use of traditional tools like MS Excel for calculations and reports is widespread despite the increasing use of business intelligence systems.

AI and big data analytics are advancing very quickly in various areas (e.g., retailing, marketing, predictive maintenance) and changing business models, organisational structures and job profiles significantly. Through the adoption of artificial intelligence (AI), management accounting too faces threats, challenges and opportuni-

Isabel Lausberg, Ruhr West University of Applied Sciences, Mülheim a.d. Ruhr,
e-mail: isabel.lausberg@hs-ruhrwest.de
Arne Eimuth, Ruhr West University of Applied Sciences, Mülheim a.d. Ruhr,
e-mail: arne.eimuth@hs-ruhrwest.de
Anne Stockem Novo, Ruhr West University of Applied Sciences, Mülheim a.d. Ruhr,
e-mail: anne.stockem-novo@hs-ruhrwest.de

https://doi.org/10.1515/9783110790320-005

ties (Hasan, 2021; Vărzaru, 2022). It is undisputed that there is potential in taking over routine tasks via Robotic Process Automation as well as to come to an improved information quality with the use of advanced and predictive analytics on a data-driven approach and AI (Losbichler & Lehner, 2021; Vărzaru, 2022). It is expected that AI applications will lead to a fundamental redefinition of existing tasks in the next years which will change the “DNA” of management accounting (Vărzaru, 2022). Big data and AI will significantly increase the efficiency and effectiveness of accounting tasks. As a result, it is expected that AI will enable management accountants to achieve considerably higher customisation of results (Weißenberger, 2021).

However, their use in management accounting and the entire finance function is restrained and lagging behind as several studies document (Casas-Arce et al., 2022; Vărzaru, 2022).

Given the advantages, what leads to the lack of use of big data and AI in management reporting? How can management reporting be transformed into state-of-the-art reporting that corresponds to current technological possibilities? What are the steps that can be taken to implement such a state-of-the art-reporting in corporate practice? What are the barriers and implementation hurdles in corporate practice to realise the potential of AI and ensure efficient and effective management reporting?

In an ongoing research project, a maturity model has been developed that defines relevant criteria for AI in management reporting and different stages on the way to an “AI-Reporting.” The model is based on the relevant literature and in-depth interviews conducted with ten highly qualified company executives and researchers with either an AI or management accounting background. Content analysis according to Mayring (2014) was used for the consolidation and evaluation of the interviews. The model is described in detail in Lausberg, Eimuth and Stockem Novo (2022). During the research project, we found that despite this path to an AI-reporting specifically designed for management reporting, practical implementation still faces barriers that prevent corporate practices from realising the potential of AI technology. While there have been numerous academic studies on such barriers to the use of AI, the authors believe that the question of the barriers to the use of AI in management reporting has not been adequately addressed. This study will help to fill the research gap by identifying and analysing various enablers and inhibitors to the use of AI in management reporting based on the conducted expert interviews and further in-depth research on these barriers. First, we describe the maturity model by Lausberg Eimuth and Stockem Novo (2022) and the relevant criteria identified here for a path to an AI-based reporting. Subsequently, we focus on the inhibitors and enablers for this development.

5.2 The Path to AI-Reporting – A Maturity Model for AI in Management Reporting

Maturity models offer a framework for developing, analysing and evaluating the capabilities of an organisation with regard to processes, products and services. Maturity models emerged out of quality management and became well-known, particularly in the software industry from the early 1990s onwards, with still increasing dissemination in other fields (Wendler, 2012). These models are characterised by identifying different stages of development based on the specifications of relevant criteria and thus indicating different levels of maturity up to the stage of completion.

The maturity model (Figure 5.1) for AI in management reporting ranges from levels 0 to 3. The levels are characterised by five dimensions, each with two to three subcriteria. Level 0 symbolises "non-algorithmic reporting" (based on Gentsch, 2019), in which no AI is used. Level 1, as "assisted reporting," assumes an isolated use of AI taking place in single parts of the reporting process. For example, AI could be used to uncover patterns in internal databases or to automate individual queries of internal

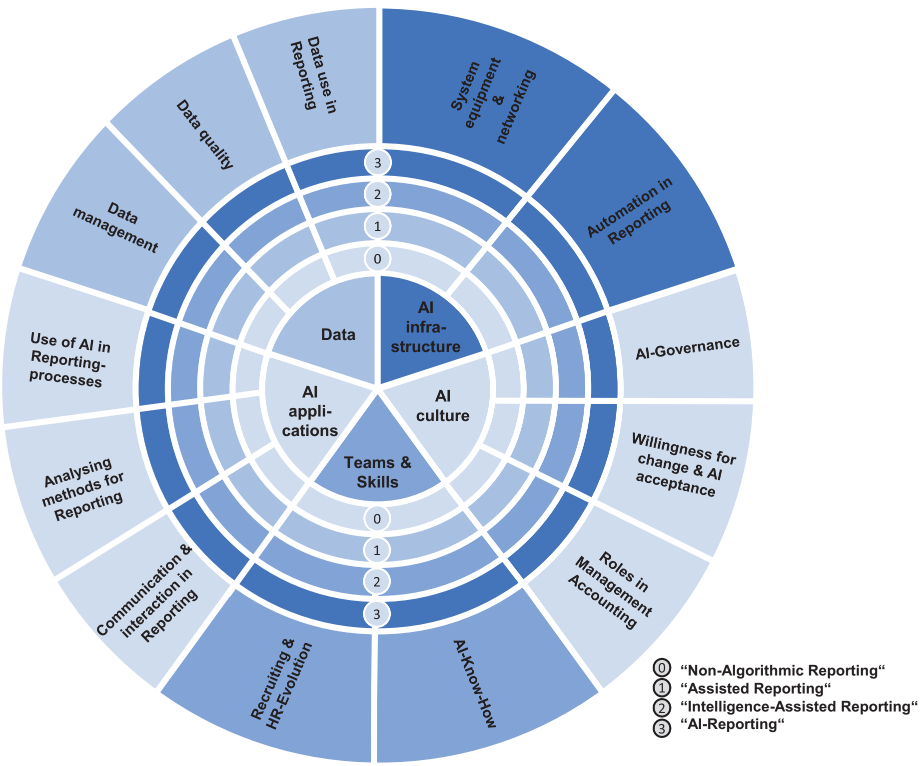


Figure 5.1: Maturity model for AI in management reporting (Lausberg, Eimuth & Stockem Novo, 2023).

data. In this stage, the AI qualification in controlling is weak and the role of controlling is largely traditional. The use of AI is much more extensive in stage 2 as "Intelligence-Assisted Reporting" and extends to essential processes in reporting. The system landscape is largely integrated, scalable and designed for big data. The staff has sound AI know-how and actively promotes the use of AI. In addition, a regulatory framework is being created that raises awareness of the quality, benefits, risks and ethical issues of AI. The maturity path finally culminates in its last stage 3 in "AI-Reporting." The highest level is characterised, among others, by a distinct and systematic data management with a comprehensive use of big data, a fully integrated and automated system landscape as well as AI-based applications in all phases and processes of reporting. Routine tasks are taken over by voice controlled and networked assistants. The handling of AI is comprehensively regulated in AI governance, with humans assuming the role of "business partners" and "pathfinders" (e.g., Langmann, 2019, p. 44).

The five dimensions and associated criteria can be described as follows.

Data: The basis of management reporting supported by AI is data. In the maturity model, the amount, variety and processing speed of data grows over the different levels through connecting and integrating various data sources. Internal accounting data are progressively supplemented, for example, by supply chain or sensor data (IoT). In higher levels, external and unstructured data such as customer comments from websites can be included, which opens up the breadth and quality of applications, for example, for business forecasts. The availability and high quality of data is crucial for a high level of development. Therefore, the acquisition, storage and processing of this data needs a systematic and well-thought-out data management. The data is thus an essential prerequisite for the creation of value through AI in management reporting.

AI infrastructure: In order to make the described data flows possible and accessible, the Dimension AI infrastructure is necessary, which at the highest level provides for an integrated and networked system landscape and automation of all reporting processes from data acquisition to report generation. The infrastructure consists of a portfolio of innovative hardware and software that enables the collection, storage, aggregation and analysis of data. In addition to a powerful ERP system, this also includes business intelligence and cloud solutions. A key success factor in the higher levels is complete integration of the systems, as isolated solutions and interfaces lead to data disruptions and considerable additional work (Keimer & Egle, 2020, p. 10). At the highest level, a homogenisation of data landscapes and systems is realised and all relevant data sources and reporting tools are integrated.

AI culture: An AI governance sets the guidelines for a safe, trustworthy and sustainable handling of AI in the company. However, to get to a higher level, not only guidelines are needed, but especially willingness to change and acceptance of AI in the company. The use of AI as a digital assistant changes the role of the management accountant, so that a redefinition of roles becomes necessary depending on the level of maturity. Increasing automation and decision support will relieve the burden of routine tasks in particular, which means that the role of the management accountant will probably change more in favour of formative and advisory tasks.

Team & skills: The increasing use of AI can only thrive if the team, skills and AI know-how grow with it. Increasing offers of AI as a service facilitate the integration of external knowledge. In the long term, internal HR development is usually more cost-efficient and sustainable, supported by appropriate recruiting measures. This is particularly true if "human oversight" is to be applied at the highest level, that is, AI algorithms are not blindly trusted but rather controlled by (appropriately qualified) humans.

AI applications: At the highest level of the maturity model, AI-based applications are feasible in all reporting processes. Data is collected, analysed and evaluated with the help of AI, can be transferred into forecasts and made available to decision-makers in a processed form. For the analysis of the data, neural networks are especially very effective for descriptive, predictive and prescriptive analyses. At the highest level, AI can be used to investigate alternative courses of action and decisions under complex requirements and restrictions. Through the possibilities of speech recognition and natural language processing, communication and interaction are changed in reporting, for example, human-machine communication can be controlled through voice commands.

5.3 Barriers of AI-Reporting

5.3.1 State of Research

Management reporting has so far been reluctant to use AI methods, concrete use cases are missing although existing and there are only a few systematic approaches (Losbichler & Lehner, 2021; Weißenberger, 2021). Applications of AI in accounting are reviewed in the literature analysis by Elmegaard, Rikhardsson and Rohde (2022) who list various research on fraud detection, auditing and several other topics. Kumar Doshi, Balasingam and Arumugam (2020) empirically examine how the use of AI creates opportunities but also threats and explore the changing role of the accountancy profession. Hasan (2021) conducts a literature review and focuses on the disruption AI causes for accountancy. He also examines the benefits and risks of AI in accounting. All of the mentioned studies focus on accounting in general. The use of AI in management reporting (as part of management accounting) remains so far a largely unexplored field (see also Casas-Arce et al., 2022).

The reasons for the reluctance of the use of AI methods are manifold. Besides human factors like resistance to change, fear of losing the job or lack of trust or knowledge, there are barriers such as insufficient data management and IT infrastructure (see also Cubric, 2020; Vărzaru, 2022). This is true not only for management reporting. Studies on organisational barriers and success factors have found a significant gap between realisation and expectations in the introduction and use of AI in various fields (Alsheibani Cheung & Messom, 2019; Ransbotham et al., 2017; Sarstedt & Wecke, 2022).

The current literature also provides multiple theoretical models that may help to explain success factors and barriers. Radhakrishnan and Chattopadhyay (2020) identify different theoretical approaches that have been used in this context in various studies such as the Technology-Organization-Environment Framework (TOE), the Diffusion of Innovation model (DOI), the Technology Acceptance Model (TAM) or the Unified Theory of Acceptance and Use of Technology (UTAUT). In the following, we will refer to the TOE, as the categorisation of the model is also used by Enholm et al. (2022), whose analysis of the most recent literature helps to dive deeper into concrete barriers and success factors.

The TOE framework identifies factors in technological, organisational and environmental contexts that affect the adoption and implementation of technological innovations. The framework has been used in various studies including for example the adoption of CRM or Software as a Service (SaaS). Enholm et al. (2022) conducted a literature analysis of AI use in organisations, including 43 papers in a thorough and systematic selection process. Through clustering, they find – in accordance with the TOE – three main categories of enablers and inhibitors that affect the "ability of organizations to deploy and utilize AI" and discuss the current state of research considering these aspects (Enholm et al., 2022). Below, we will discuss and build up on that research which is non-specific to management reporting and add the insights from our expert interviews in the context of management reporting.

5.3.2 Enablers and Inhibitors

In this chapter, we discuss the key enablers that are accelerating the adoption of AI, as well as the key inhibitors that can create barriers to AI in management reporting. As in the TOE framework, we are guided by the three categories of technical, organisational and environmental factors.

5.3.2.1 Technological Factors

Enholm et al. (2022) distinguish the technological barriers into data and technology infrastructure-related inhibitors. The maturity dimensions (see section 5.2) data, AI appliances and AI infrastructure refer to this category.

Data: AI can only be as intelligent as the data it works with. The rule is "garbage in – garbage out" and if the data quality is poor, the results are of little use. In its limited availability and often poor quality there is a main barrier on the technical side. Although internal accounting data is generally available in companies, it is not self-evident that this data is fully digital and of sufficient quality for reporting. There is not always a defined single source of truth and inconsistencies may occur due to the use of different IT systems. There is often no real-time access to the data and no integration of data beyond accounting (e.g., sensor data) and in particular of data external to the company. Furthermore, most of the existing data is historical and allows only limited forecasts for the future. Crucial for the creation of value through AI and progressing along the development path towards an AI-reporting is therefore a systematic and well-thought-out data management for the procurement, storage, maintenance and preparation of this data (Lausberg, Eimuth & Stockem Novo, 2022). One expert points out that a lot of effort is needed to train the AI and big data sets are required as training data. It is essential that these data sets are of high quality, up-to-date and reflect the variety of possible characteristics to increase the abilities, the accuracy and the reliability of the AI models (also Enholm et al., 2022).

Technical infrastructure: A major barrier to the further development of reporting lies in non-integrated systems, as isolated solutions and interfaces lead to data disruptions and considerable additional work (Keimer & Egle, 2020, p. 10). The expansion of the technical infrastructure as well as the integration of data sources and reporting tools are key factors on the way to an AI-reporting. Therefore, a homogenisation of data landscapes and systems is necessary. "First get your systems in order!: an interviewed expert demands. As AI relies on rich data sets, massive amounts of computing power, high speed and scalability are needed – whereby it is not feasible for every company to have these resources on its own; they can also be made available through service providers (Enholm et al., 2022 and the sources cited there).

5.3.2.2 Organisational Factors

Enholm et al. (2022) identify culture, top management support, organisational readiness, employees AI trust, AI strategy and compatibility as organisational barriers. The maturity dimensions AI culture and team and skills refer to this category.

Culture, top management support and organisational readiness: As AI can have a big impact on a company in terms of business model, processes and human resources, the organisation needs to be able to adapt to this change. Innovative cultures are more likely to embrace and use new technologies, and are in a better position to integrate AI into their business (Mikalef & Gupta, 2021). According to the experts interviewed, the *AI culture dimension*, which includes the readiness for change and acceptance of AI in the company, is the most important challenge for AI in management reporting (Lausberg, Eimuth & Stockem Novo, 2022). The experts note a cautious approach to technology, with trust and acceptance of technological innovation being low in many companies in Germany. Fear of change often outweighs willingness to try new things. The use of AI in reporting therefore requires cross-departmental commitment and is driven by the *top management*. That top managers play a key role in setting up this culture, as Lee et al. (2019) point out, is confirmed by several studies, finding that the support of top management is a critical factor in addressing the challenges that AI poses to organisations (e.g., Alsheibani Cheung & Messom, 2019; Alsheibani et al., 2020).

The top management support also has a crucial impact on the *organisational readiness*. Organisational readiness includes the financial resources but also the skills of the human resources (Enholm et al., 2022). As mentioned before, the technological infrastructure is a key determinant on the way to an AI-reporting. To build this infrastructure or to obtain the respective services through external service providers, a willingness by top management to allocate the needed financial resources is required. Mikalef and Gupta (2021) point out that the organisation must ensure that employees and managers have the technological knowledge, but also know which and how business functions should be used for creating value. From the experts' point of view, specialised knowledge of data interpretation or model fitting is currently lacking in most companies in Germany. The need for data scientists and IT infrastructure experts has not been fully recognised by the organisations yet. Human resources are all the more important because even at the highest level of maturity, humans are assumed to be the final authority and thus ensure "human oversight," that is, that AI algorithms are not blindly trusted but are controlled by (appropriately qualified) humans (Lausberg, Eimuth & Stockem Novo, 2022). An appropriate recruiting strategy and further training of the existing staff are recommended to overcome potential hurdles.

Employees AI trust, AI strategy and compatibility: A major barrier arises when employees have a strong resistance to change (Enholm et al., 2022). The implementation of AI is accompanied by a redefinition of the tasks of humans and technology. In order to work with AI, employees need to *trust* AI and develop a basic understanding of how AI works to be able to estimate the limitations of AI decisions. The digital disruption through AI however threatens the status quo and has a big impact on the accounting profession as one of the main professions, which AI can complement or entirely replace – questioning the long-term prospects of the profession (Kumar

Doshi, Balasingam & Arumugam, 2020). Old role models can thus represent a major barrier to AI. Newer role models for the management accountant suggest the role of the business partner or pathfinder, in which the position takes on a more advisory and formative role and is largely freed from routine activities (Langmann, 2019). This may lead to a much greater acceptance of technological progress. Schäffer and Weber (2021) emphasise that there are different ways in which AI and human intelligence can cooperate. For example, AI can create preliminary work that further undergoes a quality assurance (QA) step by humans. Major parts of management reporting, for example, analyses, can thus be automated. Alternatively, humans and AI can work together on a task, for example, on forecast models, where there are several interfaces that require human decisions, such as the choice of the database to be included or the plausibility of different scenarios. In addition to new role models, the development of guidelines for the responsible and sustainable use of AI plays an important role in reducing concerns about AI and building trust. This requires an AI governance, which also includes protecting privacy and managing risks and ethical aspects when using AI (Lausberg, Eimuth & Stockem Novo, 2022).

It is undisputed in the literature that the use of AI must be initiated and guided by a comprehensive *strategy* due to the necessary and far-reaching change AI entails. The so far shown enablers and inhibitors in Section 5.3.2 as well as their interdependencies indicate a complexity that supports the hypothesis of the current literature. The interviewed experts also agree that AI in management reporting requires an overall AI strategy and integration into the organisation (Lausberg, Eimuth & Stockem Novo, 2022).

Compatibility refers to the extent to which tasks and technology fit together. It is key to provide benefits and create acceptance. Compatibility can relate to business processes and business cases (Enholm et al., 2022; Pumplun, Tauchert & Heidt, 2019). In this context, the importance of use cases should also be mentioned, which represent a good introduction to the use of AI in management reporting and can create acceptance for AI applications. The experts' advice here is straightforward: start, try things out, gain experience and face the challenges resolutely. Use cases in management reporting can include, for example, chatbot applications, simulations, pattern recognition and process analyses with corresponding visualisation options (Lausberg, Eimuth & Stockem Novo, 2022).

5.3.2.3 Environmental Factors

Ethical and moral aspects, regulations and environmental pressure are recognised as environmental barriers (Enholm et al., 2022). These factors are mainly external factors which are usually given and non-influenceable by the company. Therefore, they are not specified in the maturity model. However, these barriers need to be recognised, their significance to be reflected upon and decisions to be made on how to deal with them.

Part of this can be achieved through an AI governance, which sets out guidelines for the company that reflect the demands of external factors. In management reporting relevant aspects are such as data protection, the detection of risks and ethical handling of data and AI as well as the corresponding responsibilities (Lausberg, Eimuth & Stockem Novo, 2022).

5.4 Conclusion

In order to leverage the evident potential of AI for management reporting, hurdles on the way to an AI-reporting need to be overcome. This article provides a first, structured overview of potential barriers on the way to an AI-reporting laid out in the maturity model of Lausberg, Eimuth and Stockem Novo (2022) and allowed to deduct first steps on how to overcome these hurdles. The comprehensive literature analysis of current research results including several meta-studies on barriers to the use of AI technology in accounting was applied to management reporting with reference to expert interviews conducted in an ongoing research project. The identified barriers were assigned to *technological* and *organisational* factors in accordance with the TOE framework and Enholm et al. (2022).

The benefits of AI as a technology depend on the input data and the methods and systems used to provide and analyse this data (*technological factors*). This also applies to management reporting: The potential of an AI-reporting results from the analysis of large and high-quality data sets, within and outside of (management) accounting, as well as historical and future-oriented data. In the analysis of the initial situation in management reporting, the data often does not correspond to the required quality: Frequently, data sets are incomplete or inconsistent with an undefined single source of truth. In addition, access to non-accounting and forward-looking data is often rudimentary (Lausberg, Eimuth & Stockem Novo, 2022). This is exacerbated by the use of disparate and rather poorly integrated systems and databases, which lead to the use of error-prone interfaces. The result is an increase in the effort required to obtain the necessary large volumes of high-quality data. This suggests, in accordance with the experts in our survey, that the system landscape needs to be operational and powerful before an AI-reporting can become economically viable.

Organisational factors, such as AI-culture, organisational readiness and top management support are also critical to the realisation of AI's potential in management reporting. AI thrives in a culture open to technology. Such an innovative culture is more likely to embrace new technologies and is in a better position to integrate AI into the management reporting process. The interviewed experts note a rather cautious approach to technology in many German companies. Lower levels of trust and acceptance of technological innovation also characterise the AI culture so that the fear of change outweighs the potential benefits. A cross-departmental commitment, driven by top management, is needed to move the culture to a more innovation-embracing environment.

Top management support also affects organisational readiness: Massive computing power, high speed and scalability are attributes of a technical infrastructure that allows AI to thrive in management reporting. This requires an appropriate level of willingness to allocate the necessary (financial) resources for this purpose by the top management. Employee skills and competencies also contribute to organisational readiness: According to the maturity model, humans will have the ultimate authority in terms of "human oversight" even at the highest level, the AI-reporting. From the expert's point of view, the majority of German companies do not currently have the necessary level of expertise in data interpretation or model fitting to make greater use of AI in management reporting. Combined with old role models in accounting, this contributes to a culture of mistrust and aversion to AI (e.g., fearing that AI will take over tasks that are currently being performed by traditional accounting roles). Experts suggest increasing training and improving recruitment strategies that help to overcome these hurdles. In addition, a greater dissemination of AI technology can be achieved by definition and acceptance of new management accounting roles which integrate, promote and enable the use of AI in management reporting.

Environmental factors that influence the adoption of AI technology are ethical and moral aspects, regulations and environmental pressure. These factors are not directly mirrored in the maturity model but may have a strong impact on how companies engage with AI. They have to be taken into account in an AI governance which defines how data protection and ethical handling of AI are assured. Moreover, these factors have to be considered in the AI strategy and its operationalisation for management reporting.

Overall, the different aspects of the TOE framework cannot be considered in isolation. They are interdependent, implying or influencing one another. The impact of interdependencies has not been addressed in this study and could support the assessment of the importance of the factors. For example, technical shortcomings may have an impact on the acceptance of AI solutions and decrease employee's trust in AI. Missing top management support will make it hard to invest in technical infrastructure or external regulations as another example of potential interdependencies may limit the exploitation of data sources.

This study provides a first analysis of potential barriers to the use of AI in management reporting within a proven framework for the adoption of new technologies (TOE). The findings obtained here, based on the exploratory nature of the expert interviews, can be tested for empirical significance in the next step. In addition, the scope of the study should be broadened, as the experience of the experts in the survey is focused on German companies.

Bibliography

Alsheibani, S. A., Cheung, D., & Messom, D. (2019). Factors inhibiting the adoption of artificial intelligence at organizational-level: A preliminary investigation. Twenty-fifth Americas Conference on Information Systems, Cancun.

Alsheibani, S. A., Cheung, Y., Messom, C., & Alhosni, M. (2020). Winning AI Strategy: Six-Steps to Create Value from Artificial Intelligence. AMCIS, 11.

Brock, J. K.-U., & von Wangenheim, F. (2019). Demystifying AI: What digital transformation leaders can teach you about realistic artificial intelligence. *California Management Review*, *61*(4), 110–134.

Casas-Arce, P., Cheng, M. M., Grabner, I., & Modell, S. (2022). Managerial accounting for decision-making and planning. *Journal of Management Accounting Research*, *34*(1), 1–7.

Cubric, M. (2020). Drivers, barriers and social considerations for AI adoption in business and management: A tertiary study. *Technology in Society*, *62*, 101257.

Elmegaard, J., Rikhardsson, P., & Rohde, C. (2022). The role of artificial intelligence in accounting – new perspectives on empirical research. http://dx.doi.org/10.2139/ssrn.4191419.

Enholm, I.M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). Artificial intelligence and business value: A literature review. *Information Systems Frontiers*, *24*, 1709–1734. https://doi.org/10.1007/s10796-021-10186-w.

Gentsch, P., (2019). Algorithmic business: Framework und reifegrad-modell. *Künstliche Intelligenz für Sales, Marketing und Service*, 43–58.

Hasan, A. R. (2021). Artificial intelligence (AI) in accounting & auditing: A literature review. *Open Journal of Business and Management*, *10*(1), 440–465.

Keimer, I., & Egle, U. (2020). Grundlagen für den erfolgreichen digitalen Wandel im Controlling. Die Digitalisierung der Controlling-Funktion: Anwendungsbeispiele aus Theorie und Praxis, 1–16.

Kumar Doshi, H. A., Balasingam, S., & Arumugam, D. (2020). Artificial intelligence as a paradoxical digital disruptor in the accounting profession: An empirical study amongst accountants. International Journal of Psychosocial Rehabilitation, 24, 873–885. https://doi.org/10.37200/IJPR/V24I2/PR200396.

Langmann, C. (2019). *Digitalisierung im Controlling*. Wiesbaden: Springer Gabler.

Lausberg, I., Eimuth, A., & Stockem Novo, A. (2022): Künstliche Intelligenz im Management Reporting – Der Weg zum „AI-Reporting“. *Controlling – Zeitschrift für erfolgsorientierte Unternehmenssteuerung*, *34*(6), 54–61.

Lee, J., Suh, T., Roy, D.,& Baucus, M. (2019). Emerging technology and business model innovation: the case of artificial intelligence. Journal of Open Innovation: Technology, Market, and Complexity, 5(3), 44.

Losbichler, H., & Lehner, O. M. (2021). Limits of artificial intelligence in controlling and the ways forward: a call for future accounting research. *Journal of Applied Accounting Research*, *22*(2), 365–382.

Mayring, P. (2014). Qualitative content analysis: theoretical foundation, basic procedures and software solution. Klagenfurt. http://nbn-resolving.de/urn:nbn:de:0168-ssoar-395173.

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, *58*(3), 103434.

Pumplun, L., Tauchert, C., & Heidt, M. (2019). A new organizational chassis for artificial intelligence-exploring organizational readiness factors.

Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). Reshaping business with artificial intelligence: Closing the gap between ambition and action. *MIT sloan management review*, *59*(1).

Radhakrishnan, J., & Chattopadhyay, M. (2020). determinants and barriers of artificial intelligence adoption – A literature review. International Working Conference on Transfer and Diffusion of IT, 89–99. https://doi.org/10.1007/978-3-030-64849-7_9.

Sarstedt, M., & Wecke, B. (2022). *Skalierung von KI im Marketing und die neue Rolle des CMO*. Science meets Practice. Wiesbaden: Springer Gabler. https://doi.org/10.1007/978-3-658-37864-6_2.

Schäffer, U., & Weber, J. (2021), Künstliche Intelligenz. *Controlling Management Review*, *65*(2), 3.

Vărzaru, A. A. (2022). Assessing artificial intelligence technology acceptance in managerial accounting. *Electronics*, *11*(14), 2256.

Weber, J., Malz, R., & Lührmann, T. (2012). *Excellence im Management-Reporting: Transparenz für die Unternehmenssteuerung*. John Wiley & Sons.

Wendler, J. (2012). The maturity of maturity model research: A systematic mapping study. *Information and Software Technology*, *54*(12), 1317–1339. https://doi.org/10.1016/j.infsof.2012.07.007.

Weißenberger, B. E. (2021). Künstliche Intelligenz als Zukunftstechnologie im Controlling. *Controlling & Management Review*, *65*(2), 8–17.

Thomas Rautenstrauch, Janis Hummel, Oliver Isoz, Simon Moser

Chapter 6
Transforming Management Accounting with Robotic Process Automation – Requirements and Implications

Abstract: This chapter explores the role and influence of robotic process automation (RPA) on the work of accountants and specifically examines the impact of RPA on the job profile, knowledge and required skills of the management accounting profession. Following a qualitative research approach, semi-structured in-depth interviews are conducted with accounting and finance experts from bigger Swiss corporations in order to complement existing literature with new empirical data from practical experience. Results show that management accountants will not be dismissed by RPA as they understand their role as internal business partners and advisors if they focus on valuable tasks and capabilities. Evidence from expert interviews and literature provides insights into a necessary change in the qualifications of the management accounting profession induced by a growing number of RPA implementations in Swiss corporations. Further studies may wish to explore the changing requirements of the management accounting job profile with extended empirical research.

Keywords: robotic process automation, RPA, automation, finance transformation, management accounting, accounting, controlling, artificial intelligence, process management

6.1 Introduction

Due to the COVID-19 pandemic and other extraordinary shocks, the relevance of lean and cost-efficient indirect processes has increased and therefore accelerated the use of digital-enabled solutions (Eklund, Kabra & Rao, 2022). Especially robotic process automation (RPA) has become a popular software technology in recent years and is one of the fastest-growing software implementations due to an increasing business de-

Thomas Rautenstrauch, Eastern Switzerland University of Applied Sciences, IFL Institute for Finance and Law, Rapperswil SG, Switzerland, e-mail: thomas.rautenstrauch@ost.ch
Janis Hummel, HWZ University of Applied Sciences in Business Administration Zurich, Center for Data Science & Technology, Zurich, Switzerland, e-mail: janis.hummel@fh-hwz.ch
Oliver Isoz, Project Manager Digital Finance, V-ZUG AG, Zug, Switzerland, e-mail: oliver.isoz@gmx.ch
Simon Moser, HWZ University of Applied Sciences in Business Administration Zurich, Center for Data Science & Technology, Zurich, Switzerland, e-mail: simon.moser@fh-hwz.ch

https://doi.org/10.1515/9783110790320-006

mand for efficiency and productivity (Gartner, 2022). With software robots being embedded in business processes, RPA can not only take over tasks of human workers but also offers massive productivity and efficiency gains by automating highly repetitive, administrative processes (Flechsig, Lohmer & Lasch, 2019).

According to Aguirre and Rodriguez (2017), processes that are suitable for RPA use are characterised by being standardised and rule-based with lower cognitive requirements, but are highly prone to human error due to manual labour as well as performed through digital data handling with access to multiple systems. Although RPA nowadays is considered to be an established technology, it is the combination of RPA and artificial intelligence (AI) that leads to various synergies beyond cost benefits through the automation of back-office processes (Lamberton, Brigo & Hoy, 2017). As a consequence, the possibilities of AI in the context of process automation brought up the scenario of a digital workforce, especially within a company's support functions (Automation Anywhere, 2022). In this context, Moffitt, Rozario and Vasarhelyi (2018) and Tripathi (2018) indicate that providers of RPA solutions often ignore the savings effects on human workforce but instead emphasise the activity shift from repetitive tasks to more creative or valuable tasks as a benefit of RPA. On the other hand, Ivančić, Suša Vugec and Bosilj Vukšić (2019) concluded that RPA has more often been implemented in various organisations in recent years than it has been investigated in the research.

Furthermore, today's capabilities for handling complex tasks, especially within back-office or support functions, have become broader and led to concepts such as *hyper automation* or *smart automation* (Lacity & Willcocks, 2016; Lasso-Rodriguez & Winkler, 2020; Schmitz, Stummer & Gerke, 2019). This also relates to accounting including business or corporate functions in a company's back-office which often have been centralised in shared service centres (SSC) during the past years (Howcroft & Richardson, 2012). Being among those corporate functions with a high estimated potential for automation, accounting is therefore expected to be an object of digital transformation that is leveraged by the increasing number of RPA implementations in SSC (Anagnoste, 2017; Lacity & Willcocks, 2016).

After years of organising back-office functions of accounting in SSC, the application of RPA combined with AI has become an inevitable trend to decrease operational costs through efficiency and accuracy (Figueiredo & Pinto, 2021). According to Lacity and Willcocks (2016), the processes and tasks in accounting that are most affected by the use of RPA include not just the classic end-to-end processes such as order-to-cash (O2C), procure-to-pay (P2P) or record-to-report (R2R) but also subprocesses like debt collection, incentive claim or trend tracking.

Despite high expectations on the various benefits of RPA mentioned in the literature, Meironke and Kuehnel (2022) point out that traceable assessments of these benefits are rare but at the same time often cited in literature without sufficient empirical evidence. Moreover, the automation of manual processes and tasks, which can currently be observed in accounting in particular, seems likely to trigger a profound change in job profiles (Harrast, 2020).

Given this background, this chapter aims to assess the current and long-term impacts of RPA robots on workplaces and job profiles in management accounting.

With regard to this overall goal, the following three research questions should be examined and answered:

RQ1: Where do resource savings through RPA show up in management accounting?

RQ2: What alternative uses are the saved resources going to?

RQ3: What new tasks and competences are emerging that will affect the work of experts in management accounting?

This chapter follows an exploratory qualitative research approach to better understand the implications of RPA for (management) accounting and is therefore structured as follows: After Section 6.1, the adopted research methodology is described, before Section 6.3 presents the findings from six interviews with accounting and finance experts of larger Swiss corporations. Finally, the chapter concludes with discussing its results and implications, reflecting on limitations and suggesting prospects for further development.

6.2 Methodology

With regards to the research objective, insights are gained from existing literature and practical experience on the use and benefits of RPA in the areas of accounting and finance. Therefore, this study was conducted using a qualitative research approach, with expert interviews serving as the main data-gathering method while being based on Grounded Theory. According to Mayring (2015), a qualitative content analysis is defined by its explicitly rule-based framework which allows for a structured and systematic handling of language material.

The interview partners were selected on the basis of a stakeholder analysis in the area of management accounting. It was found that people from the areas of consulting, financial management, HR management and accounting management were particularly suitable. This was identified by considering three main criteria for suitability: experience in the field of robotics, knowledge of the skills and responsibilities of management accountants, and sufficient experience in human resource development and planning. All three criteria were used in the evaluation with equal weighting of one-third each. After the evaluation, the stakeholders were divided into groups A, B and C based on their scores, with group A containing the most suitable and group C the least suitable stakeholders.

In total, six people were interviewed. They work in companies in the German-speaking part of Switzerland. It must be noted that the job titles mentioned in the stakeholder analysis are not always exactly transferred to the business world as functions often blur into one another. The majority of the search for suitable participants was carried out by consulting existing networks of the authors. All expert opinions were treated confidentially and anonymously. Table 6.1 summarises the expert characteristics.

Table 6.1: Overview of expert characteristics.

Expert number	Job title	Area of expertise	Date
Expert 1	Head Financial Reporting	Finance	14 June 2021
Expert 2	Lecturer in HR Management and Leadership	Consulting and HR	15 June 2021
Expert 3	Management and Founder	Consulting	24 June 2021
Expert 4	Head Finance Transformation	Finance	24 June 2021
Expert 5	Head of Controlling and Accounting	Accounting and Finance	25 June 2021
Expert 6	CFO and HR Director	Finance and HR	1 July 2021

The interviews took place between 14 June 2021 and 1 July 2021 and were conducted online, via Microsoft Teams, due to the COVID-19 pandemic. All interviews were conducted and transcribed in German because all interviewees were native German speakers. The mean duration of the interviews, excluding introductions and goodbyes, was between 23 and 36 minutes. This is at the lower limit of the recommended 30 to 45 minutes, but according to Meier, Polfer and Ulrich (2020), this could have been beneficial, as a short interview duration can lead to higher efficiency and attention. All of the six interviews were subsequently transcribed.

The main analysis of the empirically gathered data from the interviews is based on the theory of qualitative content analysis by Mayring (2015) and Gläser and Laudel (2010), using semi-standardised interview guides.

The topic blocks for the interview guide were formed based on the research questions and current literature. The first part of the interview dealt with the impact of robotics on resource planning in management accounting. The aim was to find out to what extent companies can benefit from the use of robotics. It should also be clarified whether the use of robotics leads to saving resources in management accounting. Afterwards, questions were asked about a possible shift of resources through robotics. The aim was to find out which other activities, some of them being new, could be taken over by management accountants. The third block of topics was the most comprehensive and built on the second block. The aim of this block was to find out which competencies and requirements specialists in management accounting should acquire for the activities mentioned. Before finishing the interview, an open final question was asked. According to Meier, Polfer and Ulrich (2020), this is useful in order to obtain information that is often not yet expressed but nevertheless significant for the study.

As mentioned before, the categories for the analysis of the data were deductively derived, which means that they have been derived and justified in advance from theory and the research questions. Furthermore, the structure of the data analysis has been set a priori. Mayring (2015) distinguishes between formal, content-based, typifying and scaling analytical structures, where in this case the content-based structuring was deemed to be the most suitable, as the data from the interviews could be assigned to the predefined categories.

The in-depth analysis of the interviews followed the example of Gläser and Laudel (2010) and proceeded as follows. As soon as a text passage was found in the transcriptions of the interviews that could be assigned to the corresponding category, it was marked, edited and extracted into a separate list. In a further step, the category system was reviewed and, where necessary, improved and extended. Finally, the results were presented based on the defined structuring dimension, in this case content structuring. The aim of this method was to summarise the essential statements of the interviews within the categories (Mayring, 2015). The presentation of findings follows the format of Vitharanage et al. (2020), as this table format is considered fitting for the visualisation of the main results (see Section 6.3).

6.3 Results

In the following, the main findings drawn from coding and analysing the transcribed expert interviews are summarised and visualised (see Table 6.2). After the examination of potential savings in resources due to RPA implementation, its influence on shifts in resource allocation is assessed. Lastly, a variety of possible changes in tasks and needed skill sets are defined and motivated. The defined categories of change do not only arise from the expert interviews but are also covered by existing literature.

Table 6.2: Evidence from expert interviews and current literature.

Category of change	Evidence from interviews	Evidence from literature
Savings in resources through RPA implementation		
Time savings	"The aim is to achieve efficiency gains and, above all, to create more time for other activities." (Expert 1, Position 9) "[. . .] bringing about increased automation, simply frees up time for the further development of management accounting, i.e. that the employees don't have to deal with simple, boring standard tasks, but that they can really approach more complex problems and can really contribute directly to the value creation of the company" (Expert 6, Position 9)	Kreher et al. (2020) Kokina & Blanchette (2019) Qiu & Xiao (2020) Kaya, Turkyilmaz & Birol (2019) Knauer, Nikiforow & Wagener (2020)

Table 6.2 (continued)

Category of change	**Evidence from interviews**	**Evidence from literature**
Personnel resource savings	"The number of management accountants, in itself, could decrease because the field becomes more specific or because it becomes more automated." (Expert 1, Position 62) "And the human resources that are then freed up can be used for more complex issues. This inevitably leads to a much more exciting job for the employee, if they don't just have to do the same thing over and over again, but can really take care of creative and forward-looking topics." (Expert 5, Position 5) "And now, of course, I can say, okay, if I still have a lot of manual data compilers and then I introduce Robotics, I can most likely provide the same services with fewer staff." (Expert 3, Position 12)	Kreher et al. (2020) Hauptmann et al. (2020) Qiu & Xiao (2020) Kaya, Turkyilmaz & Birol (2019)
Shifts in resource allocation through RPA implementation		
Predictive analytics	"[. . .] Your prediction for next week looks very red so that you will somehow be at index 80 instead of index 100, what measures do you have to take before this prediction materializes?" (Expert 3, Position 15) "[. . .] I think it's moving a little bit away from just backward-looking data compilation, more towards a forward-looking management accountant that really does data interpretations, that's about managing the system which gathers the data and then interpreting it accordingly." (Expert 5, Position 15)	Keimer et al. (2018) Kaya, Turkyilmaz & Birol (2019) Knauer, Nikiforow & Wagener (2020) Wolf & Heidlmayer (2019)
Process automation	"On the one hand, we use the freed-up resources to build up further automation, i.e. to further automate our processes and achieve an increase in quality." (Expert 5, Position 19)	Loitz et al. (2020) Figueiredo & Pinto (2021)
Business steering	"Another point when one looks at planning is clearly market analysis. More time for market analyses, business steering, where should the journey go and so on." (Expert 1, Position 11) "[. . .] and on the other hand, the automations also give us much more time to really take care of the analyses and the management of the company and thus to gain deeper insights." (Expert 5, Position 19)	Kümpel, Schlenkrich & Heupel (2019) Schäffer & Weber (2016) Wolf & Heidlmayer (2019)

Table 6.2 (continued)

Category of change	Evidence from interviews	Evidence from literature
Focus on value-added services	"In other words, we can automate the boring, repetitive work and then really use the freed-up resources for important things like customer contact or new projects." (Expert 5, Position 3) "And the added value now is really taking the time to understand things. So, what exactly happened? Why did it happen? And then maybe even make recommendations for action and say: Okay, in the future we have to do it this way and that way." (Expert 1, Position 11)	Kümpel, Schlenkrich & Heupel (2019) Schäffer & Weber (2016) Knauer, Nikiforow & Wagener (2020)
Business partners	"[. . .] The idea would be to automate more in this area, so that the management accountant can take more care of the business partners." (Expert 4, Position 15) "[. . .] but then to take on any advisory functions, that in turn then generates more value for the company [. . .]" (Expert 6, Position 9)	Kümpel, Schlenkrich & Heupel (2019) Rautenstrauch (2019) Keimer et al. (2018) Wolf & Heidlmayer (2019)
Data analytics	"[. . .] it's already going in the direction of something like data analyst or programmer. It's a kind of fusion of different professions." (Expert 5, Position 21) "I have the feeling that a controller will and must deal more with the technical side, in the future [. . .]. [He] must understand, what you do with it all the data, what is the goal, what do I have to achieve [. . .]." (Expert 3, Position 17)	Klein & Gräf (2020) Kümpel, Schlenkrich & Heupel (2019) Schäffer & Weber (2016) Kaya, Turkyilmaz & Birol (2019) Knauer, Nikiforow & Wagener (2020)
Changes in tasks and needed skill sets through RPA implementation		
Digital and IT expertise	"One competency that may not have been as strong as it is now were the programming skills. I think that in the future, every controller will have to acquire this knowledge in order to be able to see the possibilities and also the limitations of robotics." (Expert 5, Position 25) "[. . .] you have to understand what the possibilities are. You have to come to terms with the technologies." (Expert 3, Position 23) "[. . .] in the future, I think that the controller also has to bring along or build up the digital know-how himself. They need a certain affinity, otherwise it will be difficult." (Expert 4, Position 37)	Keimer et al. (2018) Loitz et al. (2020)

Table 6.2 (continued)

Category of change	**Evidence from interviews**	**Evidence from literature**
Methodological and analytical expertise	"[. . .] I mean something like design competence. In other words, creating models that can then be used or that are fed by robotics, AI or other options and produce reasonable outputs, i.e. forecasting models and things like that. I think that will be the future, that management accounting will also be somewhat active in terms of designing models." (Expert 5, Position 25) "I think the analytical aspect is also important. It is essential to be able to interpret the data and, above all, to draw the right conclusions, isn't it?" (Expert 6, Position 17) "Because you're really going to have the controller who, how should I say, who understands the engine, who does the data preparation, who makes sure that the data is complete and accurate." (Expert 1, Position 62)	Klein & Gräf (2020) Kümpel, Schlenkrich & Heupel (2019) Rautenstrauch (2019) Keimer et al. (2018) Wolf & Heidlmayer (2019)
Social skills	"As a controller who wants to have a real impact, I have to take a close look at the human factor. I see very, very, very big deficits there." (Expert 2, Position 25) "[. . .] And I think that's why controllers need certain soft skills, so that they can approach people and know how to get the data, the information, the relevant data from them." (Expert 6, Position 17)	Keimer et al. (2018)
Coordination and adaption-related skills	"In the future, we will have to cooperate much, much more than in the past across these supposed or actual departmental boundaries. This silo thinking is deadly." (Expert 2, Position 39) "[. . .] and the fact that not everything can be done in one place or in one person, that they [the management accountants] is strongly dependent on all data flows, on other places, on programmers, on data analysts, also a certain degree of leadership competence is necessary to be able to coordinate the people accordingly and also to be able to motivate them to work in the direction of management accounting [. . .]." (Expert 5, Position 35)	Keimer & Egle (2020) Wolf & Heidlmayer (2019)

Table 6.2 (continued)

Category of change	Evidence from interviews	Evidence from literature
Understanding processes in practice	"[. . .] the controller should have good insights into various departments, into processes. And that's where I see his tasks. If they see something that is being done well in department A, then they should take this knowledge to department B." (Expert 6, Position 17) "Theory and practice are always two worlds, of course. In practice [. . .] you have to gain experience, case experience. I would say that every deal, every contract is different. [. . .] you have to be able to apply the theory in practice." (Expert 1, Position 54)	Hermann, Stoi & Wolf (2018) Wolf & Heidlmayer (2019)
Understanding corporate structure	"I mean, if you want to upgrade, that is, provide predictive data that is relevant to business management, then you have to understand the business even more broadly." (Expert 2, Position 25)	Keimer et al. (2018) Knauer, Nikiforow & Wagener (2020)
Compliance	"What one should also, I think, keep an eye on, is the aspect of compliance [. . .]" (Expert 1, Position 68) "[. . .] what also strikes me is that a controller the longer the more also has compliance tasks, of course also according to company size [. . .]" (Expert 6, Position 21)	Kokina & Blanchette (2019) Kaya, Turkyilmaz & Birol (2019)

6.4 Conclusion

This study aims to add transparency to the existing transformation in accounting induced by RPA and to assess the current and long-term impacts of RPA on workplaces and job profiles in management accounting. Interview-derived results indicate that the use of RPA as an external driver is capable of influencing a wide range of activities in management accounting comprehensively and thus at the same time leads to a change in the previous activity profiles of these functions within organisations.

Answering the first research question relates to the scarce resource of *time*, as RPA enables software bots to take over low value-added and repetitive tasks, allowing these activities to be performed faster and with higher quality as well as accurateness.

Regarding the second research question, it can be stated that according to literature, two paths for future accountants have been identified. On the one hand, it is assumed that especially management accountants will develop in the direction of a business partner and thus become an internal consultant. On the other hand, they may specialise in the direction of a (citizen) data scientist. The results of the interviews clearly confirm these new role models, where a certain overlap between the tasks of management accountants and (citizen) data scientists exists. Therefore, it can be assumed that management accounting will increasingly support management in an advisory capacity as a business partner in strategic issues. The risk of a possible displacement of management accounting by other, possibly also external experts does not exist, especially if management accounting has extensive and specific, internal knowledge and already assumes a central role in management support. Nevertheless, (management) accountants must continuously develop their skills and knowledge in order to strengthen their supporting role in a company.

With regard to the third research question, it can be stated that the scope of management accounting is clearly changing through the use of robotics. Repetitive and manual work that offers little added value to internal customers of management accountants is automated and the time freed up is reallocated to work such as predictive and prescriptive analytics or corporate and project control. In order to be able to take over this role and carry out these tasks, various competences are necessary, especially digitalisation skills and IT knowledge, to fully exploit the possibilities of new technologies. While programming skills are not essential but beneficial for understanding the systems, social skills need to be developed, as the use of robotics puts people even closer to the centre of leadership. Besides, analytical and advisory skills will be more important in the future due to the fact that comprehensive business and market knowledge is required to advise management in the best possible way. In order to work on its own reputation among internal customers, management accountants need to communicate the added value created in the company in an adequate and understandable manner. Especially, management accountants need the ability to coordinate across and cooperate with several departments and to adapt to the increasingly agile environment.

Finally, the major conclusions of this chapter can be summarised as follows:

– Time savings are the most important contribution following the introduction of RPA in accounting, allowing accountants to free themselves from routine tasks with comparatively little added value and instead deliver greater value to their internal clients by taking on high-value services.
– With the help of RPA, classic routine accounting activities within the finance organisation are automated, such as the processing of incoming invoices or expense reports. For many companies, this automation is of great importance because bots can complete tasks faster and more accurately than humans. However, the interviews show that staff savings are not the primary motivation to use RPA in the area of accounting, as the ongoing competition for highly skilled workers obviously suggests (Yigitbasioglu, Green & Cheung, 2023).
– The job profiles of management accountants are currently changing rapidly because of increased interest in further training focusing on application knowledge in RPA, AI and other IT-related disciplines such as data analytics or process mining. In addition, this study points out that the social skills of management accountants are likely to become more important in the future.

Despite the chapter's clear and valuable implications on the impact of RPA on resource (re)allocation and required skill sets in management accounting, it faces various limitations. Although being supported by existing literature, the qualitative approach of deriving arguments from expert interviews lacks generalisability and representativeness. The validity of results is not only threatened by the small number of interviewed experts but also by heterogeneity among them regarding position and experience. Despite limiting the scope and focusing only on the influence of RPA on management accounting at Swiss corporations, general statements regarding industry or country would require further quantitative research with detailed consideration for organisational characteristics. Nonetheless, the chapter is a valuable addition to existing literature by providing important implications for understanding and handling the far-reaching influence of RPA on (management) accounting.

Bibliography

Aguirre, S., & Rodriguez, A. (2017). Automation of a business process using robotic process automation (RPA): A case study. In: J. Figueroa-García, E. López-Santana, J. Villa-Ramírez, & R. Ferro-Escobar (Eds.), *Applied Computer Sciences in Engineering. WEA 2017. Communications in Computer and Information Science* (Vol. 742, pp. 65–71). Springer. https://doi.org/10.1007/978-3-319-66963-2_7.

Anagnoste, S. (2017). Robotic automation process – the next major revolution in terms of back office operations improvement. *Proceedings of the International Conference on Business Excellence*, *11*(1), 676–686. https://doi10.1515/picbe-2017-0072.

Automation Anywhere. (2022, May 6). *Top 5 back-office functions that should be automated*. https://www.automationanywhere.com/company/blog/rpa-thought-leadership/top-five-back-office-functions-that-should-be-automated (accessed 4 January 2023).

Eklund, S., Kabra, R., & Rao, S. (2022, August 18). *Optimizing front- and back-office services in advanced electronics*. McKinsey & Company. https://www.mckinsey.com/industries/industrials-and-electronics/our-insights/optimizing-front-and-back-office-services-in-advanced-electronics (accessed 18 January 2023).

Figueiredo, A. S., & Pinto, L.-H. (2021). Robotizing shared service centres: key challenges and outcomes. *Journal of Service Theory and Practice*, *31*(1), 157–178. https://doi.org/10.1108/JSTP-06-2020-0126.

Flechsig, C., Lohmer, J., & Lasch, R. (2019). Realizing the full potential of robotic process automation through a combination with BPM. In: C. Bierwirth, T. Kirschstein, & D. Sackmann (Eds.), *Logistics Management. Lecture Notes in Logistics*. Springer. https://doi.org/10.1007/978-3-030-29821-0_8.

Gartner (2022). *Gartner says worldwide RPA software spending to reach $2.9 Billion in 2022*. https://www.gartner.com/en/newsroom/press-releases/2022-08-1-rpa-forecast-2022-2q22-press-release (accessed 19 December 2022).

Gläser, J., & Laudel, G. (2010). *Experteninterviews und qualitative Inhaltsanalyse*. VS Verlag für Sozialwissenschaften, Wiesbaden.

Harrast, S.-A. (2020). Robotic process automation in accounting systems. *Journal of Corporate Accounting & Finance*, *31*(4), 209–213. https://doi.org/10.1002/jcaf.22457.

Hauptmann, M., Denny, B., Mallon, D., Poynton, S., Schwartz, J., Van Durme, Y., Volini, E., & Yan, R. (2020). *Die soziale Organisation bei der Arbeit: Paradox in die Zukunft.* Deloitte Insights. https://www2.deloitte.com/content/dam/Deloitte/at/Documents/human-capital/at-hc-trends-2020-deutsch-kurzfassung.pdf (accessed 19 January 2023).

Hermann, K., Stoi, R., & Wolf, B. (2018). Robotic process automation in finance & controlling der MANN+HUMMEL Gruppe. *Controlling*, *30*(3), 28–34. https://doi.org/10.15358/0935-0381-2018-3-28.

Howcroft, D., & Richardson, H. (2012). The back office goes global: exploring connections and contradictions in shared service centres. *Work, Employment and Society*, *26*(1), 111–127. https://doi.org/10.1177/0950017011426309.

Ivančić, L., Suša Vugec, D., & Bosilj Vukšić, V. (2019). Robotic process automation: Systematic literature review. In: C. Di Ciccio, R. Gabryelczyk, L. García-Bañuelos, T. Hernaus, R. Hull, M. I. Stemberger, A. Ko, M. Staples (Eds.), *Business Process Management: Blockchain and Central and Eastern Europe Forum. BPM 2019. Lecture Notes in Business Information Processing* (Vol. 361, pp. 280–295). Springer. https://doi.org/10.1007/978-3-030-30429-4_19.

Kaya, C.-T., Turkyilmaz, M., & Birol, B. (2019). Impact of RPA technologies on accounting systems. *The Journal of Accounting and Finance*, *82*, 235–250. https://doi.org/10.25095/mufad.536083.

Keimer, I., & Egle, U. (2020). *Die Digitalisierung der Controlling-Funktion*. Springer Gabler, Wiesbaden. https://doi.org/10.1007/978-3-658-29196-9.

Keimer, I., Gisler, M., Bundi, M., Egle, U., Zorn, M., Kosbah, M., & Bueel, A. (2018). *Wie digital ist das Schweizer Controlling? Eine schweizweite Analyse auf Basis eines Reifegradmodells*. Hochschule Luzern – Wirtschaft. Institut für Finanzdienstleistungen Zug IFZ. https://www.hslu.ch/-/media/campus/common/files/dokumente/w/ifz/studien/wie-digital-ist-das-schweizer-controlling-hslu-deloitte-2018.pdf (accessed 7 January 2023).

Klein, A., & Gräf, J. (Eds.). (2020). *Reporting und Business Analytics*. Rudolf Haufe Verlag, Freiburg im Breisgau.

Knauer, T., Nikiforow, N., & Wagener, S. (2020). Die Bedeutung und Ausgestaltung von Robotic Process Automation (RPA) im Controlling. *Controlling*, *32*(4), 68–75. https://doi.org/10.15358/0935-0381-2020-4.

Kokina, J., & Blanchette, S. (2019). Early evidence of digital labor in accounting: Innovation with robotic process automation. *International Journal of Accounting Information Systems*, *35*. https://doi.org/10.1016/j.accinf.2019.100431.

Kreher, M., Gundel, U., Sellhorn, T., Hess, T., Köster, A., & Weiss, K. (2020). *Digitalisierung im Rechnungswesen 2020*. KPMG.

Kümpel, T., Schlenkrich, K., & Heupel, T. (2019). *Controlling & Innovation 2019*. Wiesbaden: Springer Gabler Verlag.

Lacity, M.C., & Willcocks, L.P. (2016). Robotic process automation at Telefonica O2, *MIS Quarterly Executive*, *15*(1), 21–35.

Lamberton, C., Brigo, D., & Hoy, D. (2017). Impact of robotics, RPA and AI on the insurance industry: Challenges and opportunities. *Journal of Financial Perspectives*, *4*(1).

Lasso-Rodriguez, G., & Winkler, K. (2020). Hyperautomation to fulfil jobs rather than executing tasks: the BPM manager robot vs human case, *Romanian Journal of Information Technology and Automatic Control*, *30*(3), 7–22.

Loitz, R., Bardens, A., Eiber, M., Meyer, S., & Sternisko, A. (2020, June). *Robotic Process Automation (RPA) in der DACH-Region*. PWC. https://www.pwc.de/de/rechnungslegung/robotic-process-automation-rpa-in-der-dach-region.pdf (accessed 18 January 2023).

Mayring, P. (2015). *Qualitative Inhaltsanalyse*. Weinheim: Beltz Verlag.

Meier, C., Polfer, L., & Ulrich, G.-S. (2020). *Wissenschaftsmethodik*. Zürich: Verlag SKV.

Meironke, A., & Kuehnel, S. (2022). How to Measure RPA's Benefits? A Review on Metrics, Indicators, and Evaluation Methods of RPA Benefit Assessment. *Wirtschaftsinformatik 2022 Proceedings.*

Moffitt, K., Rozario, A. and Vasarhelyi, M. (2018). Robotic process automation for auditing, *Journal of Emerging Technologies in Accounting*, *15*(1), 1–10. https://doi.org/10.2308/jeta-10589.

Qiu, Y.-L., & Xiao, G.-F. (2020). Research on cost management optimization of financial sharing center based on RPA, *Procedia Computer Science*, *166*, 115–119. https://doi.org/10.1016/j.procs.2020.02.031.

Rautenstrauch, T. (2019). *Digitale Technologien: Die Veränderungen im Controlling und Accounting*. WEKA. https://www.weka.ch/themen/finanzen-controlling/rechnungswesen/buchfuehrung/article/digitale-technologien-die-veraenderungen-im-controlling-und-accounting/ (accessed 21 January 2023).

Schäffer, U., & Weber, J. (2016). Die Digitalisierung wird das Controlling radikal verändern. *Controlling & Management Review*, *60*(6), 6–17. https://doi.org/10.007/s12176-016-0093-9.

Schmitz, M., Stummer, C., & Gerke, M. (2019). Smart automation as enabler of digitalization? A review of RPA/AI potential and barriers to its realization. In: P. Krüssel (Ed.), *Future Telco. Management for Professionals*. Cham: Springer. https://doi.org/10.1007/978-3-319-77724-5_31.

Tripathi, A.-M. (2018). *Learning Robotic Process Automation: Create Software Robots and Automate Business Processes with the Leading RPA Tool–UiPath*. Birmingham: Packt Publishing.

Vitharanage, I.-M.-D., Bandara, W., Syed, R., & Toman, D. (2020). An empirically supported conceptualisation of robotic process automation (RPA) benefits. In: *Proceedings of the 28th European Conference on Information Systems (ECIS2020).* Association for Information Systems.

Willcocks, L.-P., & Lacity, M.-C. (2016). *Service Automation Robots and the Future of Work*. Ashford, UK: SB Publishing.

Wolf, T., & Heidlmayer, M. (2019). Die Auswirkungen der Digitalisierung auf die Rolle des Controllers. In: B. Feldbauer-Durstmüller, S. Mayr (Eds.), *Controlling – Aktuelle Entwicklungen und Herausforderungen*. Springer Gabler. https://doi.org/10.1007/978-3-658-27723-9_1.

Yigitbasioglu, O., Green, P., & Cheung, M. D. (2023). Digital transformation and accountants as advisors. *Accounting, Auditing & Accountability Journal*, *36*(1), 209–227. https://doi.org/10.1108/AAAJ-02-2019-3894.

Part 4: **Organisation and Workflow**

Markus Harlacher, Andrea Altepost, Ingo Elsen, Alexander Ferrein, Adjan Hansen-Ampah, Wolfgang Merx, Sina Niehues, Stefan Schiffer, Fatemeh Nasim Shahinfar

Chapter 7
Approach for the Identification of Requirements on the Design of AI-supported Work Systems (in Problem-based Projects)

Abstract: To successfully develop and introduce concrete artificial intelligence (AI) solutions in operational practice, a comprehensive process model is being tested in the WIRKsam joint project. It is based on a methodical approach that integrates human, technical and organisational aspects and involves employees in the process. The chapter focuses on the procedure for identifying requirements for a work system that is implementing AI in problem-driven projects and for selecting appropriate AI methods. This means that the use case has already been narrowed down at the beginning of the project and must be completely defined in the following. Initially, the existing preliminary work is presented. Based on this, an overview of all procedural

Note: The competence centre WIRKsam (funding code: 02L19C600 (continuous)) is funded by the Federal Ministry of Education and Research as part of the funding measure "Regional Competence Centers of Labor Research" and is supervised by the Project Management Agency Karlsruhe (PTKA). The responsibility for this contribution lies with the authors.

Markus Harlacher, Institute of Applied Industrial Engineering and Ergonomics, Düsseldorf, e-mail: m.harlacher@wirksam.nrw
Andrea Altepost, RWTH Aachen University, Institut für Textiltechnik, Aachen, e-mail: a.altepost@wirksam.nrw
Ingo Elsen, FH Aachen University of Applied Sciences, Big Data Lab, Aachen, e-mail: elsen@fh-aachen.de
Alexander Ferrein, FH Aachen University of Applied Sciences, Mobile Autonomous Systems and Cognitive Robotics Institute (MASKOR), Aachen, e-mail: ferrein@fh-aachen.de
Adjan Hansen-Ampah, RWTH Aachen University, Institut für Textiltechnik, Aachen, e-mail: a.hansen-ampah@wirksam.nrw
Wolfgang Merx, RWTH Aachen University, Institut für Textiltechnik, Aachen, e-mail: w.merx@wirksam.nrw
Sina Niehues, Institute of Applied Industrial Engineering and Ergonomics, Düsseldorf, e-mail: s.niehues@wirksam.nrw
Stefan Schiffer, FH Aachen University of Applied Sciences, Mobile Autonomous Systems and Cognitive Robotics Institute (MASKOR), Aachen, e-mail: s.schiffer@fh-aachen.de
Fatemeh Nasim Shahinfar, Institute of Applied Industrial Engineering and Ergonomics, Düsseldorf, e-mail: f.shahinfar@wirksam.nrw

https://doi.org/10.1515/9783110790320-007

steps and methods is given. All methods are presented in detail and good practice approaches are shown. Finally, a reflection of the developed procedure based on the application in nine companies is given.

Keywords: business understanding, requirements, process model, participation, implementation of AI-systems

7.1 Introduction

Artificial intelligence (AI) offers extensive opportunities, for example, to improve knowledge management, process flows and quality outcomes. Simultaneously, AI shapes the world of work by changing, for example, tasks, requirements or skill requirements. AI acts as a tool that makes relevant information in large volumes of data accessible, picks up the structures contained therein and makes them usable (Terstegen & Jeske, 2021). The possibilities of supporting work with information technology can be used for both informational and energetic activities. Further activity components can be structured according to technical, organisational and personal focal points and enable improvements in the mental stress and strain situation of employees (Jeske, 2016). In this context, the socio-technical design of work (Altepost et al., 2021; Frost, Jeske & Terstegen, 2019) which considers humans, technology and organisation synchronously, is essential for the successful use of AI.

In order to successfully develop and introduce AI solutions in operational practice, a comprehensive process model is being tested in the WIRKsam joint project, which can be adapted in detail to the specific needs of the companies. It is intended to increase acceptance among users, to integrate the expertise of employees into the development and to promote the development of an innovative mindset. Therefore, it contains participatory elements from the very beginning of the process.

The chapter focuses on the procedure for identifying requirements for a work system that is implementing AI in problem-driven projects and for selecting appropriate AI methods. This means that the use case has already been narrowed down at the beginning of the project and must be completely defined in the following. Initially, the existing preliminary work is presented and referenced. Based on this, an overview of all procedural steps and methods is given. All methods are presented in detail and good practice approaches are shown. Finally, a reflection of the developed procedure based on the application in nine companies is given.

7.2 General Approach

7.2.1 Preliminary Work

7.2.1.1 PaGIMo/APRODI Project

Based on the theoretical foundation of Gestalt organisational consulting (Nevis, 1998), a procedure for the successful design of change processes was developed in the PaGIMo project (Thul et al., 2015) and tested as part of the APRODI joint project using digitisation projects (Bahlow et al., 2020; Gerst et al., 2021). The basic idea of this approach is that technical, organisational and human-related aspects of work system design are equally considered. The approach is based on three success factors: participation, integrity and integration. A participatory approach means that the affected employees are involved in brainstorming, planning and implementing the solution. Integrity and integration are addressed by ensuring that the digitisation solutions fit in with the corporate culture and support the employees in their work accordingly.

The APRODI methodology consists of four phases: Orientation, Focusing, Realisation and Stabilisation. These phases build on each other and are provided with feedback loops between the phases. A number of different methods are listed for the respective phases, which should be selected to suit the company (Bahlow et al., 2020).

7.2.1.2 SozioTex Project

Subject of the SozioTex project was the participatory development of an assistance system for weaving. From two phase models of system development, the phase logic and relevant specifications were adapted (Altepost et al., 2021). To map the interactions between humans, technology and organisation (HTO principle; Ulich, 1997; Ulich, 2013), the model provides for iteration loops. In this way, the possible effects of each project step in one of the three HTO areas on the other areas can be taken into account. A very central aspect is the participation of employees right from the beginning. In particular, the iterative participatory system development from mock-up to MVP is transferred to the WIRKsam process model. Like APRODI, SozioTex used a socio-technical specification sheet as the basis for system development (see Section 7.3.2.6).

7.2.1.3 Change Model by Stowasser and Suchy

The implementation of AI technologies, like all other changes in the company, requires proper change management. To address the changes and challenges caused specifically by an AI, specific AI-related change management should be applied. Therefore, we refer to the change management model of Stowasser and Suchy (2020), which deals specifically with the implementation of AI. The model pursues the idea of sensitising employees regarding the manifold changes brought about by AI applications and therefore addresses employee participation in the introduction of AI in order to increase acceptance of the technology. In order to make working with AI conducive to personality and health and to promote openness to innovation in the company, the early process-oriented involvement of employees is very important (Stowasser & Suchy, 2020).

7.2.1.4 CRISP DM

Changing a work system typically involves changes also on its technical level like changing tools or technical components. A typical use case is the replacement of a 100% manual quality control system by an AI-based decision support system that helps reducing the workload for the quality inspectors. Since its early inception, the Cross Industry Standard Process for Data Mining (CRISPDM) is the standard process for analysis, design, implementation and deployment of a production-ready solution.? (Martinez-Plumed et al., 2019; Chapman et al., 2000).

7.2.2 WIRKsam Approach

While CRISP-DM itself establishes its own phases to capture requirements for the technical part until the deployment of technical solutions, including their evaluation, it lacks support for the integration of work process improvements and focuses on the technical parts only. The APRODI process, on the other hand, was not designed for the development of technical components. Hence, there is a need to couple both processes to create the required artefacts in a well-defined way that also allows for iterative development and continuous improvement after deployment into production.

The tight coupling of both processes APRODI and CRISP-DM, for example, by interleaving them, is not advisable as this would also imply that the teams would run through the processes' phases in a synchronised fashion. Instead, we propose to couple the processes based on a set of common artefacts that are created as results from a specific process step, or process sub-step execution by joint teams (Figure 7.1). After that, the following process steps are executed unsynchronised until another common artefact must be generated or be handed over between the processes, including the artefact's approval by the creating and the receiving side. In addition to the socio-technical specifications, technical solution approaches at various stages of development are understood as artefacts. The handed-over technical artefacts are tested using an empirical and evaluative methodical approach created in the SozioTex project, leading to further iterative development of artefacts (function models) with increasing complexity and realism. Moreover, the principles of Change Model by Stowasser and Suchy (2020) are included in the process.

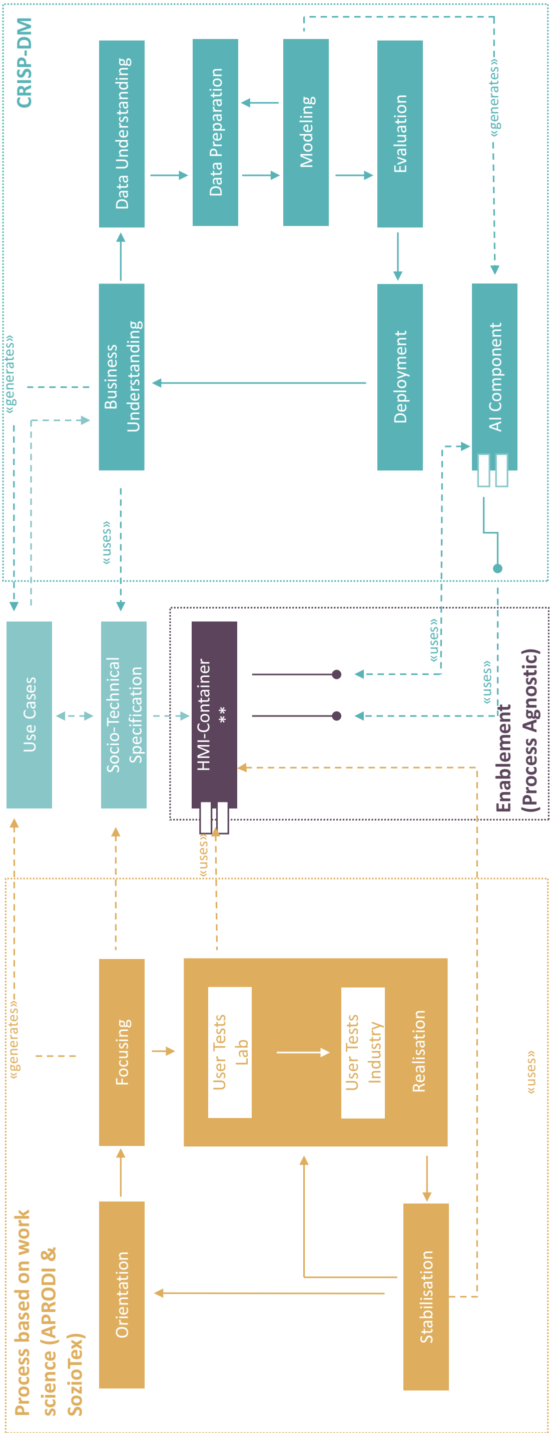


Figure 7.1: Overview of the methods for the identification of requirements on the design of AI-supported work systems in problem-based projects. (compiled by authors.)

7.3 Methods for the Identification of Requirements

7.3.1 Overview

Regarding the identification of requirements on the design of AI-supported work systems in problem-based projects, the part of focusing and business understanding has to be considered. Figure 7.2 summarises the methods and also shows a possible timeline for the application of these methods. The methods are presented in detail in Section 7.3.2.

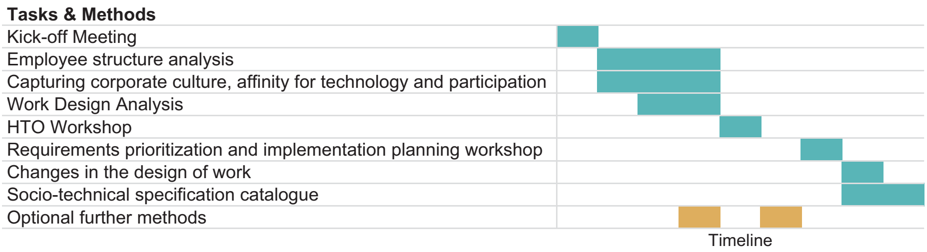


Figure 7.2: Gantt-Chart with tasks and methods for the identification of requirements on the design of AI-supported work systems in problem-based projects. (compiled by authors.)

7.3.2 Methods in Detail

7.3.2.1 Kick-Off Meeting

At the beginning, a kick-off meeting should be held in which all relevant project participants (e.g., directly affected employees and managers, company management, works council, IT officer, data protection officer and external service providers) should take part. The project objectives and the procedure with a corresponding distribution of roles should be presented and agreed upon so that all participants have the same level of information (e.g., GPM, 2019; Knechtli, 2012). A company visit and, in particular, a detailed presentation of the work system concerned or the selected pilot area also prove to be essential.

7.3.2.2 Employee Structure Analysis

The analysis of workforce structure is a fundamental part in understanding certain aspects of human resources of a company and its organisation. When looking at partner companies, analysing sociodemographic data of the workforce (age, gender, job position, education, qualification, etc.) is important firstly to know more specifics about each company. Secondly, this analysis enables the comparison of companies with each other. In addition to the aforementioned general sociodemographic data, information about the usage of digital media and periphery is requested, for example whether workers are given individual company e-mail addresses or have access to individual or shared company computers.

The data are provided by each of the partner companies through a digital spreadsheet. All data from these spreadsheets are then analysed and compared by methods of descriptive statistics. Furthermore, the data can be connected with each company's organisation chart to identify specific characteristics of each, for example possible understaffing in a department which might affect the workflow of other departments, or foreseeable understaffing because of a group of workers being close to retirement.

7.3.2.3 Capturing Corporate Culture, Affinity for Technology and Participation

A questionnaire was developed to record the corporate culture, the employees' affinity for technology and the participation practiced in the company. The questionnaire to assess the technological affinity of users of electronic devices (german: TA-EG, Karrer et al., 2009) was used to record technology affinity, and additional items were added to experience in dealing with AI-supported systems, which were already used in a questionnaire of the competence centre KOMPAKI. The assessment of organisational culture is based on the preliminary work of Martins and Terblanche (2003). Instead of the category "Behavior that promotes innovation," a category "Team and leadership" was developed. The associated items were derived based on the prior work of Martins and Terblanche (2003), Jöns, Hodapp and Weiss (2005) and Conrad et al. (2019). Another component of the questionnaire is the "PASST" instrument (Altepost et al., 2021; Wolligandt, 2019). It identifies participation methods that employees believe enable effective participation in the introduction of new technologies and that they are willing to support. The respondents assign a total of ten points to the methods they consider most promising, resulting in a ranking from which a company-specific participation concept can be developed.

The questionnaire can be filled out in both a paper-based and an online version. Based on feedback from the application companies, explanations in easy-to-understand language have also been added to make the questionnaire easier to complete, particularly in the manufacturing divisions. The results of the survey provide an objective view of the addressed topics, which is extended by a subjective component through further methods such as the kick-off or the observation. The findings make it possible to derive company-specific measures for the holistic design of work systems already in the early project phase.

7.3.2.4 Work Design Analysis

Work design analysis is pursued using two parallel, complementary approaches: On the one hand, the actual situation in the focused work system is assessed on the basis of approximately 80 items, and the expression of the item is described. The items are assigned to the categories of work content/work task, work organisation, social relations and work environment. The evaluation is initially carried out by external observers. Aspects that cannot be observed are evaluated during interviews with the employees during their work. Finally, the results are reflected on with the project management in the company in a semi-structured interview.

On the other hand, an ethnographic and more explorative approach based on participant observation is chosen: the actual situation is recorded by external observers and conspicuous features are noted. During this, the observers can also ask the observed persons about their actions, etc. The process flows are compared with existing flow charts and the documentation is checked for validity and completeness. In addition, semi-structured individual interviews with the observed person(s) are conducted to classify the observations and to ask questions specific to the application, yet outside of the observation. The results of both approaches are synthesised and form the basis for the HTO workshop.

7.3.2.5 HTO Workshop

The core idea of the HTO (human-related, technical and organisational) workshop is to bring the expertise and needs of the employees into the redesign of the work system at an early stage and to align them with the company's goals (from the business understanding) and the HTO aspects. Along the actual work process, which does not always correspond to the company's target processes, it is discussed with employees, researchers from occupational science, computer science and textile technology, as well as enablers, for example, which functions the AI should and can take over, how the information flows and responsibilities run at the workplace as well as in the work system as a whole, which interactions with other process steps and participants must be taken into account, and which organisational changes are necessary and possible. In order to map this appropriately for processes with varying degrees of structure, these can be documented using the K3 method (Killich et al., 1999; Nielen, 2014). Through this detailed joint tracing of the work activity in focus, the researchers develop an understanding for the concerns of the employees, but also for the goals and, if necessary, limitations of the research partners of the respective other disciplines. Conversely, the scope of the targeted solution becomes clear to the employees and how their expertise and wishes are incorporated there. Indications of employees' fears that could lead to acceptance problems, as well as initial findings on qualification needs, are also desirable and valuable results. The HTO workshop thus results in concrete requirements for the socio-technical specifications (Section 7.3.2.8). In this way, a socio-technical system can emerge in which humans and technology jointly contribute to the output of the work system (e.g., Hughes, 1993; Weyer, 2008).

7.3.2.6 Requirements Prioritisation and Implementation Planning Workshop

All approaches considered in preliminary works assume a common set of requirements at the beginning of the first operational process iteration. CRISP-DM puts the creation of the respective artefacts into the phase "Business Understanding" (BU), which includes the goal setting of the project, the definition of requirements, assumptions, constraints and the financial and operational planning.

CRISP-DM's Business Understanding phase focuses, as the overall process, on the creation of technical artefacts. AI-supported work systems often require additional components. Hence, the original setting of the BU phase must be supplemented with additional tasks that are executed in this phase to generate the additional information needed. The typical setting here is a set of workshops that involve domain experts, work process experts, data science/AI experts and expertise from the IT departments that shall operate the technical part of the solution. The artefacts that emit from these workshops must be approved by all parties. The change-management process that is later applied, for example if requirements must be adapted, has to involve all the mentioned parties to perform an exhaustive impact and risk analysis if the change is applied.

7.3.2.7 Changes in the Design of Work

Based on the requirements and goals defined in the workshop detailed before, the items already used in Section 7.3.2.4 are used to assess which changes can be expected as a result of the introduction of the planned AI system. The guideline developed for this purpose first provides for an assessment of whether a change will result with regard to the respective work system elements as a result of the measures associated with the introduction of AI. An expected change is to be described here. This is followed by an evaluation and description of the expected target state. Analogous to the evaluation of the current state, it is necessary to evaluate and describe whether the item can be regarded as rather applicable or rather not applicable. Based on this analysis, relevant open questions are derived with regard to work design, which are discussed in a workshop with knowledge holders and decision-makers. This is intended to sharpen the understanding of the future work system and to determine possible, accompanying work design measures.

7.3.2.8 Socio-technical Specification Catalogue

To develop a digital system – here in the context of AI – concrete requirements are needed that have to be implemented in programming. This is often a technology-oriented process. However, following the HTO principle requires that person- and organisation-related aspects of the work system should also be designed simultaneously and that these aspects should also flow into the development of the technical system. The requirements specification usual in systems development must therefore be supplemented accordingly so that a socio-technical system can emerge as a result. The requirements for the specifications can be determined using various methods, but always involve the affected employees contributing their expertise, requirements and needs. Roth and Kötter (2020) refer to a survey as well as observational interviews in the context of the work activity, which also include the use of opportunities of the technology and the addressing of possible risks – for example, “curtailment of degrees of freedom of the user” (Roth & Kötter, 2020). They distinguish between regulatory and task-, organisation- and technology-related requirements (Roth & Kötter, 2020). Gloy et al. (2017) and Altepost et al. (2021) also chose a multistage, action-oriented approach. Group discussions and expert interviews first produced a catalogue of requirements for personal, technical, organisational and legal goals (Gloy et al., 2017). These were incorporated into the specifications, as was the subsequent workplace and process analysis, including observation interviews (Altepost et al., 2021; Lemm, 2016). In WIRKsam, too, the requirements from the business understanding and HTO workshop are transferred into a socio-technical specification per use case. It not only serves to develop the technical system but also takes up information for work and organisational design in accordance with the HTO principle.

7.3.2.9 Optional Further Methods

Since it is not necessarily possible to identify and quantify all relevant aspects for holistic system design in the short observation time or workshops conducted, it may be necessary to apply additional methods as an option. These could include self-recordings (e.g., of process steps, types of errors occurring or proportions of time), multimoment frequency analyses or user tests during the introduction of new IT-supported systems.

7.4 Lessons Learned and Summary

Especially for the development of AI applications, it is essential to define the requirements for the AI system quickly in the early project phases in order to be able to enter into the activities that prepare the development of the AI model, such as data understanding and data preparation. Moreover, frontend development should start early, to have the possibility of evaluating and improving the human–technology interaction.

This requires an approach that efficiently, yet completely, captures the work design requirements. The latter requires not only that the managers are taken into account as decision-makers in the requirements analysis, but also that, in particular, the affected employees, as later users of the assistance systems, are included in the recording of the requirements. Such an approach could be developed in the WIRKsam project and, in particular, the individual methods could be tested. A further evaluation will take place in the further course of the project at up to 6 application companies still to be identified. In these companies, methods can also be tested that are aimed at the (strategic) selection of a use case.

By applying the methods outlined above, it was possible to develop a comprehensive understanding of the problem and a solution approach for the perspective design of the work system. The participatory workshop formats met with a positive response, as system design was discussed from a wide variety of perspectives. In addition, it turned out that work processes do not always run as they appear in manuals. However, these actual processes are essential for digital support and, if necessary, for reflecting on the working methods practiced to date. The participation of employees is therefore indispensable, not only for reasons of later acceptance but also because of their expertise.

It proves advantageous to demonstrate the expected gain of knowledge and benefits of the methods for the application companies and their involved employees at an early stage. Future activities should determine the relevance of the methods – especially the methods of the quantitative coverage of organisational and cultural frameworks – for the tailored design of the introduction process.

Bibliography

Altepost, A., Merx, W., Rezaey, A., & Löhrer, M. (2021). Neue Soziotechnische Systeme in der Textilbranche. Mensch-Technik-Interaktion und Ergebnisse der partizipativen soziotechnischen Systemgestaltung für ein Assistenzsystem in der deutschen Textilindustrie. Aachen: Institut für Textiltechnik und Institut für Soziologie der RWTH Aachen University (Hrsg.). https://doi.org/10.2314/KXP:1758273097

Bahlow, J., Bartels, E., Beck, I., Bendel, A., Held, G., Hennemann, J., Kötter, W., Latniak, E., Lennings, F., Müller, F., Roth, S., Schlink, B., Schmid, J., Terstegen, S., Thönis, M., & Vollborth, T. (2020). Arbeits- und prozessorientiert digitalisieren – Vorgehensweisen, Methoden und Erfahrungen aus dem Projekt APRODI. https://www.arbeitswissenschaft.net/angebote-produkte/broschueren/ue-bro-aprodi-praxisbroschuere-2020.

Conrad, R.W., Weber, M.-A., Lennings, F., & Würfels, M. (2019). 360°-Analyse – Lean-Management-Methoden nachhaltig nutzen. Anwendung am Beispiel der 5S-Methode. Ifaa – Institut für angewandte Arbeitswissenschaft. https://www.arbeitswissenschaft.net/newsroom/pressemeldung/

news/360-analyse-lean-management-methoden-nachhaltig-nutzen-anwendung-am-beispiel-der-5s-methode.

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. *SPSS inc*, *9*(13), 1–73.

Frost, M., Jeske, T., & Terstegen, S. (2019). Die Zukunft der Arbeit mit Künstlicher Intelligenz gestalten. *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, *114*(6), 359–363. https://doi.org/10.3139/104.112106.

Gerst, D., Köpp, T., Bartels, E., Bendel, A., Held, G., Kötter, W., Latniak, E., & Lennings, F. (2021). Betriebliche Digitalisierung erfolgreich gestalten – Sozialpartnerschaftliche Orientierung für ein partizipatives soziotechnisches Vorgehen. https://www.arbeitswissenschaft.net/angebote-produkte/broschueren/ue-bro-aprodi-broschuere-sozialpartner-2021.

Gloy, Y.S., Lemm, J., Löhrer, M., Kerpen, D., Ziesen, N., & Saggiomo, M. (2017). *Anforderungskatalog für Assistenzsysteme 4.0 an Textilmaschinen am Beispiel einer Webmaschine*. Aachen: Shaker.

GPM. (2019). *Kompetenzbasiertes Projektmanagement (PM4) – Handbuch für Praxis und Weiterbildung im Projektmanagement in zwei Bänden*. Nürnberg: GPM Deutsche Gesellschaft für Projektmanagement.

Hughes, T. P. (1993). *Networks of Power – Electrification in Western Society, 1880–1930*. Baltimore: Johns Hopkins University Press.

Jeske, T. (2016). Digitalisierung und Industrie 4.0. *Leistung & Entgelt* Nr 2.2016, 3–46.

Jöns, I., Hodapp, M., & Weiss, K. (2005). Kurzskala zur Erfassung der Unternehmenskultur. https://doi.org/10.23668/psycharchives.11386.

Karrer, K., Glaser, C., Clemens, C., & Bruder, C. (2009). Technikaffinität erfassen–der Fragebogen TA-EG. *Der Mensch im Mittelpunkt technischer Systeme*, *8*(2009), 196–201.

Killich, S., Luczak, H., Schlick, C., Weissenbach, M., Wiedenmaier, S., & Ziegler, J. (1999). Task modelling for cooperative work. *Behaviour & Information Technology*, *18*(5), 325–338. https://doi.org/10.1080/014492999118913.

Knechtli, H. (2012). Ohne Kick-off-Meeting kein Erfolg. *Unternehmer Zeitung*, *2012*(7/8), 36–37. https://doi.org/10.26041/fhnw-2855.

Lemm, J (Ed.) (2016). *Empirische Ergebnisse einer Arbeitsprozess- und Arbeitsplatzanalyse zur diversitätsgerechten Assistenzsystem-Entwicklung in der Textilbranche*. Aachen: Shaker.

Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez Orallo, J., Kull, M., & Lachiche, N. (2019). CRISP-DM twenty years later: From data mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering*, *33*(8), 3048–3061. Doi: 10.1109/TKDE.2019.2962680

Martins, E. C., & Terblanche, F. (2003). Building organisational culture that stimulates creativity and innovation. *European Journal of Innovation Management*, *6*(1), 64–74. https://doi.org/10.1108/14601060310456337.

Nevis, C. (1998). *Organisationberatung – Ein gestalttherapeuthischer Ansatz*. Köln: Edition Humanistische Psychologie.

Nielen, A. (2014). Systematik für die leistungs- und zuverlässigkeitsorientierte Modellierung von Arbeitsprozessen mit kontrollflussorientierten Notationssystemen. Reihe Industrial Engineering and Ergonomics Band 15 [Doctoral dissertation, RWTH Aachen]. Shaker.

Roth, S., & Kötter, W. (2020). Soziotechnisches Lastenheft: Herangehensweise bei der Anforderungsermittlung an digitale Assistenzsysteme am Beispiel eines mittelständisch geprägten Industrieunternehmens. Gesellschaft für Arbeitswissenschaft e. V. (Hrsg.): Digitaler Wandel, digitale Arbeit, digitaler Mensch? Bericht zum 66. Kongress der Gesellschaft für Arbeitswissenschaft. Dortmund: GfA-Press.

Stowasser, S., & Suchy, O. (Eds.) (2020). Einführung von KI-Systemen in Unternehmen. Gestaltungsansätze für das Change-Management. [White paper] München: Plattform Lernende Systeme.

Terstegen, S., & Jeske, T. (2021). Digitalisierung und Künstliche Intelligenz nutzen – Chancen und Anforderungen der Arbeitsgestaltung. *Arbeitsmedizin Sozialmedizin Umweltmedizin*, *56*(1), 12–14.

Thul, M.J., Longmuß, J., Zink, K.J., Steimle, U., Lange, K., & Kötter, W. (2015). Modelle für einen erfolgreichen Veränderungsprozess. In: Zink, K., Kötter, W., Longmuß, J., & Thul, M. (Eds.), *Veränderungsprozesse erfolgreich gestalten*. VDI-Buch. Berlin, Heidelberg: Springer Vieweg. https://doi.org/10.1007/978-3-662-44702-4_6.

Ulich, E. (1997) Mensch-Technik-Organisation: ein europäisches Produktionskonzept. In: Strohm, O., & Ulich, E. (Eds.), *Unternehmen arbeitspsychologisch bewerten: ein Mehr-Ebenen-Ansatz unter besonderer Berücksichtigung von Mensch, Technik und Organisation* (Vol. 10, pp. 21–38). vdf Hochschulverlag AG.

Ulich, E. (2013). Arbeitssysteme als soziotechnische Systeme–eine Erinnerung. *Psychologie des Alltagshandelns*, *6*(1), 4–12.

Weyer, J. (2008). *Techniksozoiologie – Genese, Gestaltung und Steuerung soziotechnischer Systeme*. Weinheim, München: Juventa.

Wolligandt, C. (2019). Partizipation an der Digitalisierung – Erarbeitung und Evaluation eines partizipativen Konzepts zur Implementierung eines Assistenzsystems in der deutschen Textilindustrie. [Unpublished Master Thesis]. Institut für Soziologie der RWTH Aachen University.

Nicolai Krüger, Agnis Stibe, Jacqueline Krüger

Chapter 8
Plug and Play AI – How Companies Can Benefit from AI as a Service

Abstract: The vast possibilities of artificial intelligence (AI) have recently opened to a broad user base, for example, via ChatGPT. Entrepreneurs and innovators are increasingly developing AI-based products, and companies can solve problems with a growing range of available AI as a Service (AIaaS) products. However, there is only limited research available in the field of applied AI for business practices so far, maybe due to the novelty of the topic. We aim to build an initial step for theory in this field by putting AIaaS into the spotlight and forming the intersection of applied AI and its practical wildlife as AIaaS. We aim to contribute an empirically driven taxonomy of *Plug and Play AI* tools, a method common in information systems. Thus, we contribute to the theoretical foundation of AIaaS for further research and address practitioners with practical guidance in the broad AI application space.

Keywords: artificial intelligence, AI, NoCode, LowCode, ChatGPT, AI as a Service

8.1 Introduction

The recent development of artificial intelligence (AI) models and applications has faced many publicly known milestones during the last months. OpenAI's *Generative Pre-trained Transformer (GPT*[1]*)* combined the existing language model with an easily

1 See http://chat.openai.com (accessed 30 January 2023).

Acknowledgement: We thank OpenAI for the free research access to OpenAI playground and ChatGPT (15 December 2022 Version).

Note: The views expressed herein are those of the authors and do not necessarily reflect the views of affiliated institutions.

Nicolai Krüger, IU International University of Applied Science, Information Systems, Düsseldorf, e-mail: nicolai.krueger@iu.org
Agnis Stibe, Department of Business Innovation, The Business School, RMIT University Vietnam, Ho Chi Minh City, Vietnam; Department of Informatics, Faculty of Engineering, Built Environment and Information Technology, University of Pretoria, Pretoria, South Africa; INTERACT Research Unit, University of Oulu, Oulu, Finland, e-mail: agnis@agnisstibe.com
Jacqueline Krüger, University of Osnabrück, Institute for Empirical Economic Research, Osnabrück, e-mail: jacque.krueger@gmail.com

https://doi.org/10.1515/9783110790320-008

accessible and usable chat interface and might have defined a new level of user experience (UX) in terms of human–machine interaction with AI. This also applies to *DALL-E*,[2] a deep learning model also developed by OpenAI for creating digital images based on natural language descriptions, which offers a much better UX compared to existing alternatives like *Midjourney*[3] (currently only available on Discord) or *Stable Diffusion*[4] (currently only available for people with the hardware capability and knowledge to install and run machine learning models). Local startups like *neuroflash*[5] built their product pipelines on top of such technologies and benefit from well-documented application programme interfaces (APIs) and transparent fees to form their own value proposition for their target audience. This illustrates how AI technologies face a major change: From technological, computer scientific, or engineering-driven systems to broadly accessible applications. Further, interfaces to use these AI technologies have changed radically. While access to and usage of AI models has been limited to programmers with certain experience in AI-oriented languages like Python for quite some time, nowadays, many applications are ubiquitous and require zero or very limited coding experience. The recent hype around apps like *Lensa*[6] is illustrative for this shift, where social media has been flooded with AI-generated cartoon-like portraits. While these consumer-related applications gain much attention in public media, the academic community is facing an ongoing dialogue about the technological impact of language generation tools for scholars. Besides, the startup ecosystem is also facing major developments, as products can be designed with AI, built upon AI or sold with the help of AI-generated media content. Of course, this might also apply to existing business models. However, many AI models for enterprises and large companies were already existing before the above-described recent trend of AIaaS applications. Thus, the chapter focuses on the innovation and startup-related sector and the implementation of AI beyond classical programming. We are going to refer to this kind of AI under the umbrella term *Plug and Play AI*, which we are suggesting throughout this chapter as easily accessible AI (see also our extended working definition in Section 8.2). While AI model development and intensive programming (*Pro-Code*) is a field of interest for rather technical scientific areas, we want to bring *NoCode* and *LowCode* into the spotlight of our research contribution and form the intersection of economically relevant but technically less complex implementations of AI (Desmond et al., 2022). Therefore, we formulate the following research questions:

(1) How can technologies, applications and code fragments under the umbrella of *Plug and Play AI* be clustered?
(2) How can a taxonomy of *Plug and Play AI* look like?

2 See https://openai.com/dall-e-2/ (accessed 30 January 2023).
3 See https://midjourney.com/ (accessed 30 January 2023).
4 See https://huggingface.co/spaces/stabilityai/stable-diffusion (accessed 30 January 2023).
5 See https://neuroflash.com/ (accessed 30 January 2023).
6 See https://prisma-ai.com/lensa (accessed 30 January 2023).

The chapter is structured as follows: We first provide an overview of the relevant literature in Section 8.2. The research design is presented in Section 8.3, followed by the results of the data gathering and interpretation process in Section 8.4. We discuss our findings and derive implications for research and practitioners in Section 8.5 and conclude in Section 8.6.

8.2 Background

The rise of AIaaS represents a process towards the diffusion and ubiquity of applied AI for companies of all sectors and sizes (Sundberg & Holmström, 2022). For instance, modern banking and finance are intensively driven by outsourced IT services (Bhatia, 2022), including AIaaS. Another example is automation in marketing and sales, which includes creative, conceptual and semantic tasks (Gipp, 2021) and has become widely accessible through AIaaS for a broad range of industries. Even teaching and training, an area commonly known as an entirely human-focused one, has become a market for AIaaS providers, synthesising professional trainers, academic lecturers and professors (Pandey, Mishra & Tiwari, 2021; Synthesia, 2022). Nevertheless, the phenomenon of AIaaS and its relevance for both entrepreneurship and for grown companies and organisations have been underestimated. Potentially due to the interdisciplinary character of AIaaS: While computer science usually acts on a much deeper (technical) level, economics and management research would traditionally throw economic measures onto AI.

Taxonomy development for novel technologies is a fundamental contribution to the field of information systems, computer science and related disciplines. A major instructional study has been presented by Nickerson, Varshney and Muntermann (2013) and taken as a reference for taxonomic research projects since then. However, slight improvements and derivatives have occurred since then, including more practical recommendations for rigour and accurate taxonomy research designs (Kundisch et al., 2021). By far, not every new toolchain requires a taxonomy. Typically, infographics fill this gap for rather practical viewpoints nowadays, clustering applications in a visual form. Though, in the field of academia, taxonomies may help scholars to gain a basic understanding of a new learning field, and researchers can build their own research upon structure-providing contributions of others. For instance, Barn and Barn (2016) presented a cybercrime taxonomy, which they also thrived forward to an interrelated model of the discovered objects, basically generating an ontology as a result. Lempert and Pflaum (2019) created an Internet of Things (IoT) reference architecture based on their explorative IoT taxonomy development. For *Plug and Play AI*, we deem a taxonomy as beneficial notably because of its growing practical relevance and due to its novel character as a technology, further shaping future trends.

Another building block of our research idea is the way software is created (see Figure 8.1). In traditional computer science classes, we teach high- and low-level programming languages, programming paradigms and software architecture. In this chapter, we will refer to this paradigm as *ProCode*. However, recent years have brought broader acceptance of the so-called *citizen developer*, a rather visual and model-based approach of software design (Sahay et al., 2020). We believe that this trend is groundbreaking for the research purpose of the chapter: *Plug and Play AI* is easily attachable to existing data and infrastructure through such novel rapid development tools. To fasten up standard procedures, like CRUD database interactions (CRUD: create, read, update, delete) in combination with a simple frontend and sometimes individual code fragments, *LowCode* came up. Still managed under the responsibility of the IT department, developers with limited experience in certain technologies might be able to create business applications much faster than from scratch. Finally, *NoCode* forms a new kind of application development, significantly driven by three aspects: First, its UX-driven and visual appearance follow a state-of-the-art look and feel and makes it thus, second, usable for *citizen developers* with narrow technical knowledge. This is based on the gamified nature of *NoCode* tools, for instance, through direct visual feedback when a data connection between a *NoCode* app and a database has been established. Third, connectors of the so-called integration platform as a service (*iPaaS*) environment form a cloud-only ecosystem with easy-to-use APIs, making data linkage possible on scale.

For our work, we will combine *NoCode* and *LowCode* as *Plug and Play*, due to its easily applicable nature. Thus, our extended working definition for *Plug and Play AI*

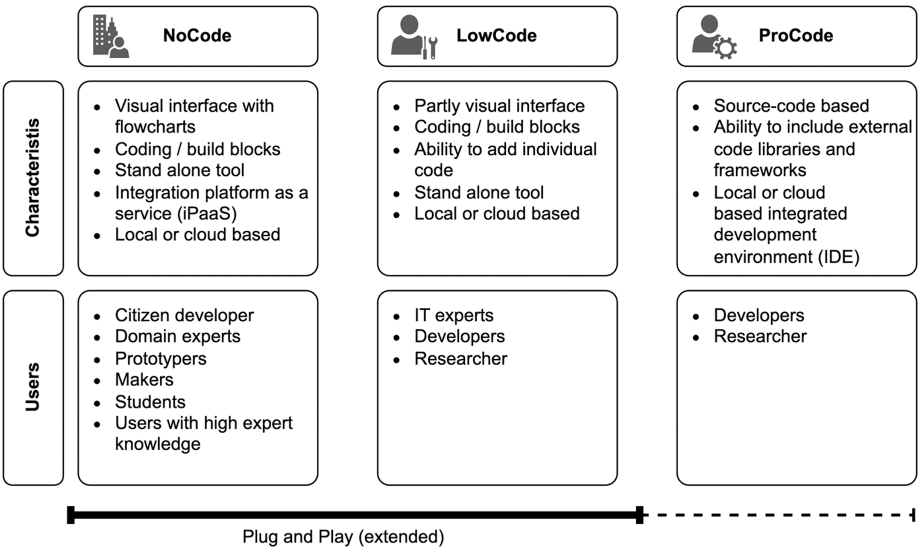


Figure 8.1: Characteristics and users for NoCode, LowCode and ProCode (adapted from Di Ruscio et al., 2022 and Desmond et al., 2022).

can be summarised as: An application (stand-alone or modularised), utilising AI methods (such as *Deep Learning*, *Neural Nets*, etc.), which is accessible and/or re-useable (in other software products) either without any (*NoCode*), with almost zero custom additional code (*LowCode*), in contrast to ProCode, where experienced coding skills are a prerequisite. In contrast to AIaaS, Plug and Play AI is not necessarily based on cloud infrastructure (Alla, 2020).

8.3 Research Design

We chose taxonomy design as an explorative paradigm for our research because the development and accessibility of such a novel technology (see Section 8.1) can hardly be accumulated into existing frameworks, definitions and research artefacts. Classification and taxonomy building for novel research fields is an important contribution for researchers and practitioners and the groundwork for later concepts to build upon, for example, semantic web technologies applying AI such as knowledge graphs (Van Rees, 2003). Taxonomy design is a well-elaborated and well-described method within the information system methodical landscape. The main concept, which has also been applied to this study, is based on an iterative data processing approach, including meta-characteristics and ending conditions, finally leading to a taxonomical hierarchy (Nickerson, Varshney & Muntermann, 2013).

As technological taxonomies are mainly driven by practice and industry, practical resources need to be taken as field data (Lempert & Pflaum, 2019). In order to find promising data sourcing of potential *No-* and *LowCode* AI firms, we first focused on the *German AI Association (KI Verband)* and accessed its membership companies (n = 390). Capterra (n = 50) and Producthunt (n = 43) were further, international, resources for curated lists of AI tools which we included, as well as seven other publicly accessible web resources (see Table 8.1). In total, different data gathering and crawling techniques delivered 553 potential companies or tools (referred to as suspects, including a title and a URL), whereof 504 went into our further data processing (the 49 remaining elements could have been eliminated immediately based on a first cross-check of title, headlines, etc.).

As a next step, we applied 15 filters to reduce our data to the defined scope (see Table 8.2). We eliminated AI tools that would not fit under our extended *Plug and Play AI* scope as defined in Section 8.2. As inclusion criteria, we defined that there should be a dedicated AI use case description available, to ensure that AI is the core of the respective tool in a transparent way. We also included only publicly available apps. In addition, we selected tools that are easy to use and access, in the sense that no specific IT knowledge should be required to use them. Further, tools with Business-to-Customer (B2C) focus were excluded, as this chapter focuses on AIaaS, thus only considering Business-to-Business (B2B) or Business-to-Business-to-Customer-relations (B2B2C). Finally,

Table 8.1: Discovered input sources. (compiled by authors.)

Source name	Suspects	Data gathering approach	Results
Geekflare	8	Manual access	10
KI-Bibliothek	6	Manual access	6
Capterra	50	AI-powered scraping with ScrapeStorm	49
KielAI	14	Manual access	8
KI Toolparty	29	Manual access (public content only, due to paywall)	27
VisualMakers	1	Manual access (dublette)	0
wemakefuture	12	Manual access	1
KI Verband	390	Python crawler	388
AI for All	43	Manual access	15
Total	**553**		**504**

we excluded tools that are not mentioned on a functioning website, as we deem this as a basic layer for validity of a tool. Lastly, we excluded duplicates. This resulted in a final list of 284 products which we then included into the further coding process.

Table 8.2: Applied filter criteria. (compiled by authors.)

Selection criteria (example)	Rule for exclusion (example)
Dedicated AI use case description (Exemplary description: "We use Random Forest Analysis to predict xyz . . .")	Obviously non-AI, like regular tools, algorithms, etc. (Calendly was mentioned in some articles as AI tool, but it is obviously just a fancy booking application)
	Workflow automation tools (i.e., n8n is a tool for cloud automation and often mentioned when it comes to AI, but actually it is a visual API connectivity platform)
	Only claiming AI for Search Engine Optimisation or marketing purposes (i.e., "AI-powered" or "AI-embedded" without actual AI usage)
Public app	Tailormade AI/model (Exemplary description: "Together with an external software vendor, we built an AI model to . . .")
Easy to use and access (limited IT knowledge and data management skills required)	AI frameworks for advanced AI experts (i.e., binaries, repositories, etc. which are only executable by experienced AI developers with extensive statistical knowledge)
B2B or B2B2C	B2C or fun products with no practical value (i.e., bodymaker is a personal sleep-tracking AI, or Lensa, an app that turns photos into art)
Website up and running	Tool not accessible through search engine / cannot be found

Next, we enriched the data set with two additional sources: First, we used OpenAI's API (text-davinci-003) to deliver us – where possible – a one-sentence description of each tool in our data set. The source code and the AI-generated descriptions are provided in our research data repository.[7] While we had to try out different prompts, we learned that the quality and density of the answers varied significantly. Then, we used ChatGPT-3 (updated December 2022) for the same task and ran into technical restrictions quickly (limited input and output due to the non-commercial version used). However, some answers looked useful to us, and others (including our implemented termination commands) got ignored by ChatGPT. Thus, we decided to use the AI-generated description wherever possible, but only in triangulation with further data to apply cross-checks, as described as follows. To reduce the manual processing time, we implemented a web-crawling snippet to download the HTML title and meta-description from the URLs in our dataset. We found this step being helpful due to the high trust and informative value of the additional data, which is – presumably – generated by the companies themselves. We provide the complete data set on GitHub. While we used MaxQDA to build our final hierarchical coding structure, we made our research data accessible as a spreadsheet to maximise compatibility.

We chose a batch size of n = 100 per iteration for the coding process. According to Kundisch et al. (2022), the first coding iteration on our data followed an explorative nature, resulting in an initial coding set with the main coding classes:

- *Code-Level*: Describing the extent of understanding, writing or processing source code required to use the application.
- *AI-Level*: Categorising the technical characteristic of the data point, in terms of rather application-like or more abstract variants.
- *Users*: Coding class to describe different user types.
- *AI Use Case*: High-level category of the use case.

After the fourth iteration, we found that no further changes in our coding structure (including sub-classes) would be providing any additional value. This fits to the suggested ending criteria in literature (Nickerson, Varshney & Muntermann, 2013). We applied some further data validity checks within our research team to strengthen the objectivity of our results via independent checks and following discussions, but without further adaptations of our final results.

7 Our research data repository can be found via https://github.com/nicolaikrueger/plugandplayai/.

8.4 Results

First, we present the hierarchical order of classes and subclasses in Table 8.3. The full hierarchical tree can be accessed in our research data set, as we plotted only those subclasses with more than ten occurrences (see Table 8.3). We have found three types of AI levels, five types of users and accordingly roles in terms of their AI interaction and a wide range of Use Cases, which encompass sectors, such as *Marketing & Sales* or *Health, Pharma, Biotech*, as well as core features the respective products are providing on a cross-sectional basis, such as *Generative AI (Text)* or *Automation*. We also found ProCode tools in our dataset, which was unexpected due to our initial filtering. Those ProCode products can be classified as easily applicable but as part of the development and maintenance chain, for example, addressing users with cloud infrastructure management skills. As those products are somewhat in-between Plug and Play AI and ProCode, we decided to keep those tools as part of our results.

Table 8.3: Taxonomical classes, subclasses and frequency. (compiled by authors.)

Coding class	Subclass	Frequency*
Code-Level	NoCode	190
	LowCode	48
	ProCode	46
AI-Level	AI-Tool	156
	AI-Powered	69
	AI-Model	59
User	Domain Expert	136
	AI Scientist	66
	Anyone	51
	AI Operator	26
	Maker	5
AI Use Case	Rapid AI Model Development & Deployment	59
	Marketing & Sales	33
	Chatbot Platform	26
	Natural Language Understanding	21
	Generative AI (Text)	18
	Health, Pharma, Biotech	14
	Automation	13
	Manufacturing & Engineering	13
	Business Analytics	11

*Sub-classes with frequencies <10 have been excluded from this overview.

We further interpreted our data under the lens of user perception (see Table 8.4). Our taxonomy differentiates between five users according to different AI roles (*AI Scientist, AI Operator*) and business cases (*Anyone, Domain Expert, Maker*), respectively.

For a better understanding, in Table 8.4 we added one example per user which we have randomly chosen from our dataset.

Table 8.4: Users and AI roles in the field of Plug and Play AI. (compiled by authors.)

User	Description and goal	Example from the data
Domain Expert	Expert in a certain field, branch or occasion with the goal to apply AI in this field.	Mindpeak is a NoCode AI-Tool for Pathologists.
AI Scientist	Expert or scientist for AI as such with the goal to build, implement or maintain AI services.	Prenode is a decentralisation infrastructure model for AI scientists.
AI Operator	Prompt engineer, intensive AI user or admin for AI with the goal to run AI-driven tasks.	ChatGPT is an AI tool for generating text.
Maker	Product specialist with the goal to build a product upon (existing) AI tools.	QnA Maker is a chatbot platform which takes existing FAQ-HTML structure as input.
Anyone	Professional of any background with the goal to ease job tasks with the help of AI.	DeepL is an AI tool for translation and proofreading.

The AI levels defined in the taxonomy are *AI-Tools*, *AI-Powered* and *AI-Model*. While *AI-Tools* encompasses products which support users for a single task, *AI-Powered* refers to AI products which cover more complex tasks driven by AI technology. *AI-Model* represents the most advanced level. While we started our research under the perspective that we might find a clear match between AI roles and *Plug and Play* tools in terms of accessibility, applicability and so forth, the results are rather counterintuitive (see Table 8.5): Only the AI role *Maker* is not covered by model-based solutions in our dataset, all other combinations can be found. This could be explained by the business specificities required for product creation. For instance, anyone can train their own AI model using Google's *Teachable Machine*. Nevertheless, it is not suitable for building another product upon it (e.g., the *Maker* role), as it follows a more educational and explorative nature. Also, this phenomenon could be induced by the specific sources of our dataset. Another reason could be that AI products for *Makers* might have been excluded during our data selection process, as we excluded tools requiring advanced AI skills (see Section 8.3).

Table 8.5: Mapping AI roles and tools. (compiled by authors.)

	AI Operator	AI Scientist	Anyone	Domain Expert	Maker	Total
AI-Model	9	38	6	6	0	59
AI-Powered	2	16	34	13	4	69
AI-Tool	15	12	11	117	1	156
Total	26	66	51	136	5	284

Focusing on the contingency table (see Table 8.6) of code levels and use cases, it becomes evident that for our data in scope, most tools have been classified as *NoCode*. *LowCode* as well as the included subset of *ProCode* are quite underrepresented, with almost close to zero tools per use case, except for the rather technical use case *Rapid AI Model Development & Deployment.* In this specific use case, most tools are *LowCode* and even a substantial number of *ProCode* tools could be identified. For most use cases in our data set, apparently no coding skills are required at all.

Table 8.6: Mapping code level and use cases*. (compiled by authors.)

	NoCode	LowCode	ProCode	Total
Automation	10	2	1	13
Business Analytics	8	2	1	11
Chatbot Platform	25	1	0	26
Generative AI (Text)	16	1	1	18
Health, Pharma, Biotech	8	2	4	14
Manufacturing & Engineering	9	1	3	13
Marketing & Sales	30	2	1	33
Rapid AI Model Development & Deployment	15	27	17	59

*Use cases with a total occurrence of <10 have been excluded from this overview for relevance reasons.

8.5 Discussion and Outlook

Interdisciplinary research on AIaaS is still rare. The recent debate about ChatGPT and other generative AI tools has intensified the discussion, relevance and knowledge seeking in this research area. First, this chapter contributes to this field by offering a more practical term and definition for AIaaS: *Plug and Play AI*. Under this umbrella term, for which we also provide some background and differentiation based on the literature, we include *NoCode* and *LowCode* solutions with AI-character and also *ProCode*, though the latter to a limited extent. Thus, the field of *NoCode* and *LowCode* approaches can also benefit from the findings of this chapter. We believe the provided taxonomy in Table 8.3 is beneficial for researchers as it represents a structured first theoretical contribution in the rather new field of AIaaS. Furthermore, we want to encourage researchers to continue an interdisciplinary discussion and research in this field and therefore, we make our data set public. The implications for practitioners are as follows: IT decision-makers can benefit from this chapter by the provided landscape of tools we collected for generating the taxonomy. While our data set is limited by the fact, that it does not cover all existing tools in this rapidly developing market, it can nevertheless help to find valid selection criteria for decision-makers.

We believe that we can contribute to the benefit of other scholars, academics and practitioners by providing data and code.

8.6 Conclusion and Outlook

The aim of this chapter is to reframe the academic discourse of AI under the lens of recent AI applications, like ChatGPT, which represents a new, ubiquitous and easily applicable category of AI tools or elements. We approached this phenomenon with an explorative view based on existing tools in the field, conducting an empirical analysis. We provide a taxonomy to build a first theoretical understanding of AIaaS, which can help practitioners, for example, during technology and startup scouting, but also support scholars for a better understanding of the interdisciplinary nature of such AI applications. In future research, we aim to thrive our data forward towards an ontology to explore underlying AI models, code libraries and tools. This might be achieved on the basis of interpreting code dependencies on GitHub, at least for those applications listed in this chapter, which are open source.

Bibliography

Alla, D. (2020). Artificial Intelligence on Information Services. https://doi.org/10.2139/ssrn.3737164.

Barn, R., & Barn, B.S. (2016). An ontological representation of a taxonomy for cybercrime. 24th European Conference on Information Systems (ECIS 2016), Istanbul, Turkey.

Bhatia, M. (2022). *Banking 4.0: The Industrialised Bank of Tomorrow*. Singapore: Springer Singapore.

Desmond, M., Duesterwald, E., Isahagian, V., & Muthusamy, V. (2022). A No-Code Low-Code Paradigm for Authoring Business Automations Using Natural Language. arXiv preprint arXiv:2207.10648.

Di Ruscio, D., Kolovos, D., de Lara, J., Pierantonio, A., Tisi, M., & Wimmer, M. (2022). Low-code development and model-driven engineering: Two sides of the same coin? *Software and Systems Modeling*, *21*(2), 437–446. https://doi.org/10.1007/s10270-021-00970-2.

Gipp, M. (2021). Wachstumstreiber Marketing Automation. *Marketing und Sales Automation*: *Grundlagen–Umsetzung–Anwendungen*, 33–46. https://doi.org/10.1007/978-3-658-21688-7_3.

Kundisch, D., Muntermann, J., Oberländer, A. M., Rau, D., Röglinger, M., Schoormann, T., & Szopinski, D. (2022). An update for taxonomy designers. *Business & Information Systems Engineering*, *64*(4), 421–439. https://doi.org/10.1007/s12599-021-00723-x.

Lempert, S., & Pflaum, A. (2019). Vergleichbarkeit der Funktionalität von IoT-Software-Plattformen durch deren einheitliche Beschreibung in Form einer Taxonomie und Referenzarchitektur. *HMD Praxis der Wirtschaftsinformatik*, *56*(6), 1178–1203. https://doi.org/10.1365/s40702-019-00562-1.

Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, 22(3), 336–359.

Pandey, C. K., Mishra, V. K., & Tiwari, N. K. (2021). Deepfakes: When to Use It. In: 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART) (80–84). IEEE. https://doi.org/10.1109/SMART52563.2021.9676297.

Sahay, A., Indamutsa, A., Di Ruscio, D., & Pierantonio, A. (2020). Supporting the understanding and comparison of low-code development platforms. https://doi.org/10.1109/SEAA51224.2020.00036.

Sundberg, L. T., & Holmström, J. (2022). Towards "Lightweight" artificial intelligence: A typology of AI service platforms. *AMCIS 2022 Proceedings*. 13. https://aisel.aisnet.org/amcis2022/sig_odis/sig_odis/13.

Synthesia. (2022). Create professional training videos in 60+ languages. https://www.synthesia.io/learning-and-development (accessed 4 November 2022).

Van Rees, R. (2003). Clarity in the usage of the terms ontology, taxonomy and classification. *CIB Report*, *284*(432), 1–8.

Part 5: **HR and Employment**

Timm Eichenberg, Nils Pudill, Britta Rüschoff, Anne Stockem Novo, Michael Vogelsang

Chapter 9
Developing Personas of Ideal-type Candidates in AI-related Jobs

An Exploratory Study Based on the Analysis of Online Job Postings

Abstract: This chapter presents a data-driven exploratory approach for identifying ideal-type candidate personas in the field of artificial intelligence (AI) based on the analysis of online job postings. The data used for analysis was collected from online job platforms for the German labour market. Latent Dirichlet Allocation (LDA) was used for topic modelling. A meaningful LDA solution with four topics was obtained, which formed the basis for identifying four different ideal-type personas that are sought for in AI-related job postings. The chapter provides an interpretation of the personas and their skill profiles, expertise and accountability level. The results can help job seekers, recruiters, and policymakers to better understand the skills and experiences that are sought by employers in the field of AI.

Keywords: personas, artificial intelligence, online job postings, profiles, skill, tasks

9.1 Introduction

Artificial intelligence (AI) changes business models, organisations and jobs. After years of research and development, today the diffusion of AI is driven by economic parameters. Workers who mainly perform routine tasks were the first to be affected by computerisation (Autor, Levy & Murnane, 2003). This assessment may change as generative AI affects the jobs of journalists, translators and artists. Also, the tasks of

Acknowledgements: We thank Tom Staab for his valuable contributions in writing and applying the Python code for the analysis.

Timm Eichenberg, Hochschule Weserbergland, Hameln, e-mail: eichenberg@hsw-hameln.de
Nils Pudill, Ruhr West University of Applied Sciences, Mülheim a.d.R.,
e-mail: nils.pudill@hs-ruhrwest.de
Britta Rüschoff, FOM Hochschule für Oekonomie und Management, Düsseldorf/Münster,
e-mail: britta.rueschoff@fom.de
Anne Stockem Novo, Ruhr West University of Applied Sciences, Mülheim a.d.R.,
e-mail: anne.stockem-novo@hs-ruhrwest.de
Michael Vogelsang, Ruhr West University of Applied Sciences, Mülheim a.d.R.,
e-mail: michael.vogelsang@hs-ruhrwest.de

https://doi.org/10.1515/9783110790320-009

computer scientists will change as code is becoming a commodity which can be produced automatically.

Businesses are developing and implementing AI technology to support their activities or make AI technology the core of their business model itself. As a result, companies' labour demands in the field of AI are increasing. Therefore, we are interested in the profiles and skills which are looked for in the field of AI. First, we use a data-driven approach for identifying candidate personas in the field of AI based on online job postings. Second, we provide an interpretation of the personas and their skill profiles, expertise and accountability level. The results can help job seekers, recruiters, and policymakers to better understand the skills and experiences that are sought by employers in the field of AI.

9.2 Skills and Jobs Research

9.2.1 Skill-Biased Technological Change

AI is one of the most recent examples for technological change. On a macroeconomic level, the consequences of technological change are discussed intensively. Skill-biased technological change (SBTC) models point out that technological change increases the productivity and hence demand of skilled workers. The approach postulates that the skill premium, the wage gap between skilled and unskilled workers, increases depending on the supply of skilled workers and the elasticity of substitution between skilled and unskilled work. However, the empirical evidence between skill-biased technological change and wage inequality was "surprisingly weak" in the past (Card & DiNardo, 2002), p. 776). Subsequently, the focus has shifted to the differences between routine and nonroutine tasks (Autor, Levy & Murnane, 2003) and the polarisation thesis that the employment shares for high-paid persons and for low-paid personal service workers are growing. Overviews are provided for example by Hornstein, Krusell and Violante (2005) and Aghion et al. (2022).

Most recently, Acemoglu and Restrepo (2022) expanded the classical SBTC model by a task-based approach. The elasticity of substitution between skilled and unskilled labour becomes endogenous and depends on the characteristics of new tasks. They also take into consideration AI. They interpret AI as a technological platform that can be used for automation or leads to new labour-complementary tasks. As new tasks can foster skilled or unskilled work, the skill premium also depends on the type of new tasks created by AI (Acemoglu & Restrepo, 2022). Empirically, Acemoglu et al. (2022) show that the AI exposures of establishments are associated with negative and positive changes in the skill set required.

9.2.2 AI Jobs Research

Several authors argue that AI jobs require general or soft skills, such as problem-solving, creativity and teamwork (Squicciarini & Nachtigall, 2021) or social and emotional skills (Samek, Squicciarini & Cammeraat, 2021). The content of AI-related jobs is hence becoming more diverse (Verma, Lamsai & Verma, 2022). Samek, Squicciarini and Cammeraat (2021) further argue that there are complementarities between different occupational categories (managers, AI professionals, technicians) as recruitment of those often happens in the same organisation at the same time. Today, business knowledge is as important as technical skills for working in business intelligence and big data projects (Debortoli, Müller & vom Brocke, 2014), which is further supported by the emergence of business clusters such as "Data Science & Engineering," "Software Engineering & Development" and "Business Development and Sales" (Anton, Behne & Teuteberg, 2020).

9.2.3 Methodological Approaches to the Analysis of AI-related Job Postings

Online job advertisements have been used as empirical databases at least since the year 2000, when Koong, Lui and Lui (2002) identified and ranked job postings in the field of information technology. Data are derived from several online platforms such as monster.com (Debortoli, Müller & vom Brocke, 2014) and hotjobs.com (Koong, Lui & Lui, 2002), indeed.com (Verma, Lamsai & Verma, 2022) or aggregators such as Burning Glass Technologies (BGT) (Acemoglu et al., 2022; Squicciarini & Nachtigall, 2021). When analysing and clustering the data, a broad range of statistical and specific Natural Language Processing (NLP) methods such as k-means (Litecky et al., 2010), Latent Semantic Analysis (LSA) (Debortoli, Müller & vom Brocke, 2014) or matrix factorisation (Anton, Behne & Teuteberg, 2020) is used. Ternikov (2022) provides a more recent overview of the methodology employed in IT job advertisement analysis and points to the fact that the approach used varies with the sample size of the data available.

In this chapter, Latent Dirichlet Allocation (LDA) is used, an approach which is applied for topic modelling. LDA is also used by Gurcan and Cagiltay (2019) who identify 48 skill topics, and by De Mauro et al. (2018), who determine nine skill sets and map them to four "Big data job families."

9.3 Methodology

This study uses the concept of personas to comprehensively summarise the results generated from the job postings regarding the aspired candidates' professional profiles along with the tasks and activities of the advertised jobs. The concept of a persona originated in human–computer interface design with the intent to facilitate user-centred designs by generating typical user profiles (Cooper, 1999). In its original application, personas were used to guide design processes by providing designers with evidence-based lifelike characters of users generated from available data on the core features (e.g., needs, goals, limitations, predicted behaviour) of the user. Since then, the use of the concept has expanded to areas outside human–computer interface design and today, a persona often generally refers to a fictional yet lifelike character that bundles the core features of, for example, a typical user (Nielsen, 2019), customer (Lehnert, Goupil & Brand, 2021) or, in this case, the targeted candidate. The approach taken in this study hence deviates from the original goal of the concept as a point of reference for design processes. Instead, we adopt the idea of a persona as a representation to identify the typical aspired candidate for a job in the area of AI, whose features do not necessarily have to be united in its entirety in one particular person.

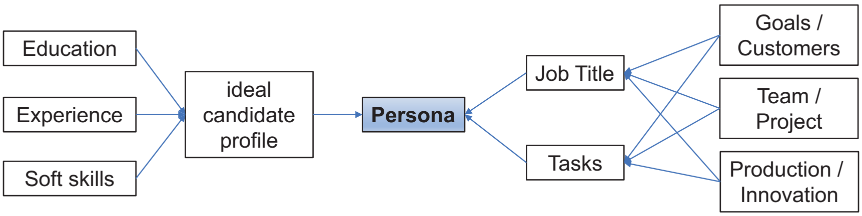


Figure 9.1: Building blocks of personas.

Figure 9.1 illustrates the structure of our approach: A persona is defined by characteristics of the aspired candidate (left side of Figure 9.1) and by characteristics of the advertised job (right side of Figure 9.1). In terms of aspired candidate characteristics, defining criteria are education, experience and soft skills. These aspired candidate characteristics are mapped to the job title and the tasks described in the job postings which are determined by the goals and customer orientation of advertising employer, its internal organisation and team structures, as well as innovation and production processes. With this persona-oriented approach, we expand the primarily skill-based view of related research and apply a more holistic view on the applicants targeted in online job postings.

9.4 Empirical Analysis

9.4.1 Data

The data subjected to analysis have been collected from three online job platforms for the German labour market from March to September 2022. The search term was restricted to "Künstliche Intelligenz" [engl.: Artificial Intelligence], yielding results for job postings with the keywords "Artificial Intelligence" in German and English language.

The following pre-processing steps were conducted: (1) Checking for duplicates by applying a cosine similarity measure and deleting duplicate entries, yielding in 8152 jobs postings. (2) Separating the job postings into job title and job profile, including degrees, and tasks. Depending on the data source, the separation of data could take place by considering structural elements such as HTML tags or by using an ontology approach with keywords. The keywords were manually updated by data exploration. (3) Normalisation of the data, converting to lowercase and elimination of stop words. (4) Partial language translation, depending on the model applied. (5) Stemming of the words.

9.4.2 Model Selection

Following our exploratory research approach, we focus on unsupervised models of machine learning to identify clusters of job postings. For the model selection, we first assumed that a set of (ideal-type candidates) profiles and a set of tasks in the clusters identified are related to each other unambiguously. Under this assumption, we used a hierarchical cluster approach and matching criteria based on the frequency distribution in a contingency table. k-means clustering with $k = 2$ was employed in three loops, so that we received a tree structure of clusters in three levels. We used TF-IDF vectors as inputs for k-means clustering as well as a Sentence BERT-Networks (SBERT) transformer called distiluse-base-multilingual pretrained model (Reimers & Gurevych, 2019). Fine tuning based on the corpus has been applied. With the multilingual SBERT embeddings, the language of the job postings should not matter. Interestingly, the results were similar to the first k-means loop with untranslated data. This suggests that in job postings written in English, specific tech-orientated skills are requested. In the next iterations, the matching criteria did not provide meaningful results, since the tasks in the job posting are too homogeneous.

We therefore deny the assumption of strongly separated clusters and use LDA for topic modelling (Blei, Ng & Jordan, 2003). The LDA algorithm assumes that words are characteristic for topics and topics are characteristic for documents. However, the topics cannot be observed directly in the documents, therefore they are latent. In this chapter, we use the LDA approach for classification of profiles, assuming that each latent topic represents an ideal-type persona. This assumption is in line with the former analysis, showing that profiles in the job postings are more heterogeneous than the tasks.

The topics themselves are described by a vector of words. The conceptual advantage of LDA over classification algorithms is that it allows for a distribution of words over the topics, for example a word can be part of more than one topic. Theoretically, the best allocation of the words to topics and topics to documents is found by maximising a joint probability function. A restriction of LDA is the assumption that the words are independent from each other. So the order of words does not influence the results ("bag-of-words assumption" (Blei, 2012)).

In this chapter, we apply the LDA function of the Gensim package for Python, which uses the variational inference approach for optimisation (Hoffman, Blei & Bach, 2010). This algorithm is very fast at finding a stable solution. Pivotal for the LDA algorithm is the parameter "number of topics." We used an expert rating method to determine the most meaningful number of topics. We started with a solution containing a high number of topics (e.g., seven) and reduced the number of topics until a solution with a meaningful differentiation between the topics was reached. The LDA run with four topics delivered the most plausible solution. This is confirmed by the coherence scores. Generally, "coherence" of a topic model follows the definition that a set of statements is coherent if the facts within the statements support each other (Röder & Both, 2015). In this chapter, we used three different measuring approaches: UCI (Newman et al., 2010), UMASS (Mimno et al., 2011) and NPMI (Bouma, 2009). The first two are named after co-authors from the University of California and University of Massachusetts, respectively, and the third abbreviation stands for Normalised Pointwise Mutual Information. To calculate scores, we used implementations also provided by the Gensim topic modelling framework. The coherence scores C(K) for K topics show a steadily declining trend by nature. Using the "elbow-method" considering the inflection points of C(K), the UCI and UMASS approaches confirm the expert's choice of four topics for the model (see Figure 9.2).

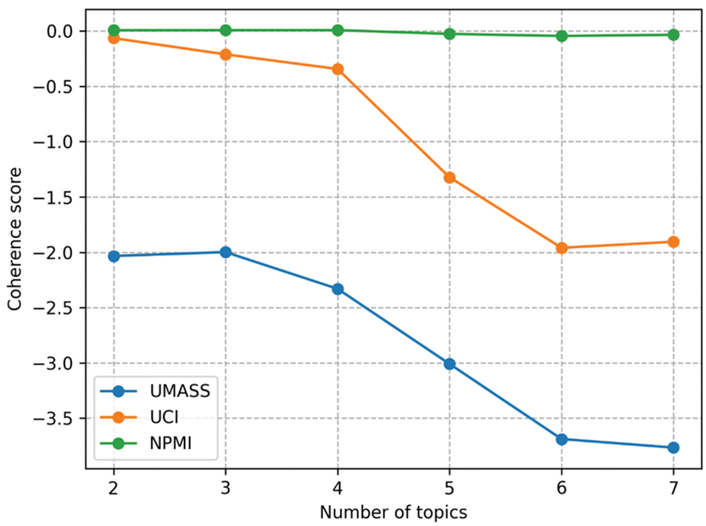


Figure 9.2: Coherence scores.

After topic modelling, we determined which postings are related to a topic and identified the tasks and job titles of these postings to describe the personas. The results are presented in the next section. Finally, we checked the results for robustness.

9.5 Results and Interpretation

9.5.1 Interpretation

Considering the model's output for four topics, characteristic keywords and the topics' positions in the coordinate system could be analysed. From the results, four different ideal-typical personas were derived, with each topic forming the basis for one persona.

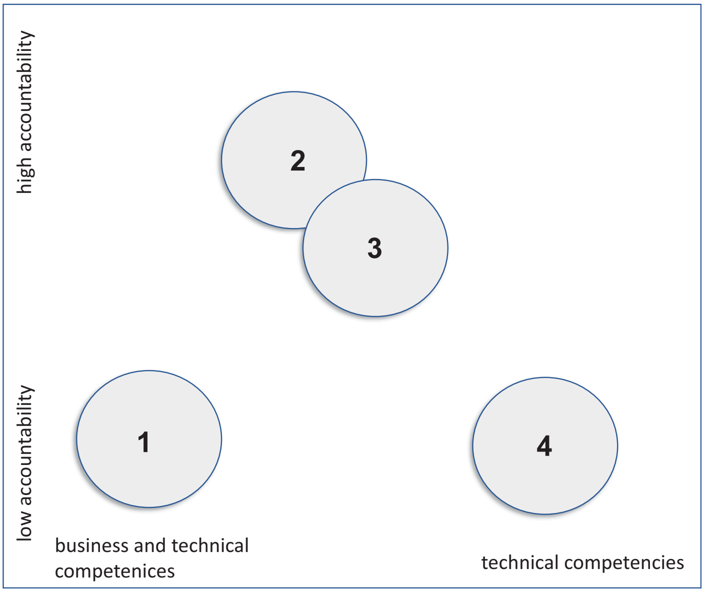


Figure 9.3: LDA model solution for four topics.

Figure 9.3 shows the stylised topics, based on the visualisation from the Python library pyLDAvis. The visualisation on two axes is to some extent arbitrary since the distribution of the topics in the coordinate system is multidimensional. In pyLDAvis, the axes (or principal components) are not specified beforehand but are generated based on a probability distribution of the words that are indicative of a certain topic. Therefore, Figure 9.3 serves entirely illustrative purposes. The meaning of the axes is likewise subject to interpretation and can only be construed conjointly with the interpretation of the four personas.

The horizontal axis can be interpreted as the level of expert competencies (left side = business and technical competencies; right side = technical competencies only). Business competencies are related to an education in business and economics [German: “Wirtschaftswissenschaften”] and/or managerial experience. The vertical axis can best be interpreted as overall accountability for results (bottom = low; top = high). Accountability refers to the person or team who is ultimately accountable for the success or failure of the outcome and who possesses the decision-making authority to approve or disapprove of the work completed by the responsible persons or teams.

9.5.2 Interpretation of the Four Personas

The proposed personas can be regarded as an explanatory approach for ideal-typical applicants in demand in the German labour market in the field of AI. In order to differentiate the four personas more efficiently, short designations for each persona were chosen. Each designation is based on the interpretation of the topics and commonly used generic IT job titles. The designations of the personas follow the interpretation of the results and do not necessarily have to be used in exactly the same wording in current or future job postings. The word clouds in this chapter show the words with the highest frequency per topic but not the most salient words. Therefore, the word clouds for the profiles show less differentiation than the LDA depicted.

9.5.2.1 Persona 1 – Junior Project Member

Persona 1 in the lower left corner of Figure 9.3 represents a candidate who is ideal for entry-level positions. Experiences [German: “Erfahrungen”] often refers to be “first experiences,” as became clear when analysing all sentences in the profiles containing the German word “Erfahrungen.” The fields of study of this persona can be business and economics [German: “Wirtschaftswissenschaften”] as well as computer sciences [German: “Informatik]. The interpretation of the tasks suggests involvement in development projects, mostly on a supporting level. We give this persona the designation “Junior Project Member.” The word clouds in figure 9.4 provide an illustration of the most frequent keywords for persona 1.

9.5.2.2 Persona 2 – Senior AI Manager

The second persona, located in the upper left corner of Figure 9.3, takes responsibility [German: “Verantwortung”] for customers. The desired educational background can be in business and economics or computer sciences. On average, the required work experience is longer than for persona 1 (which, however, cannot be seen in the word cloud). We label this persona as “Senior AI Manager.” The word clouds in figure 9.5 provide an illustration of the most frequent keywords for persona 2.

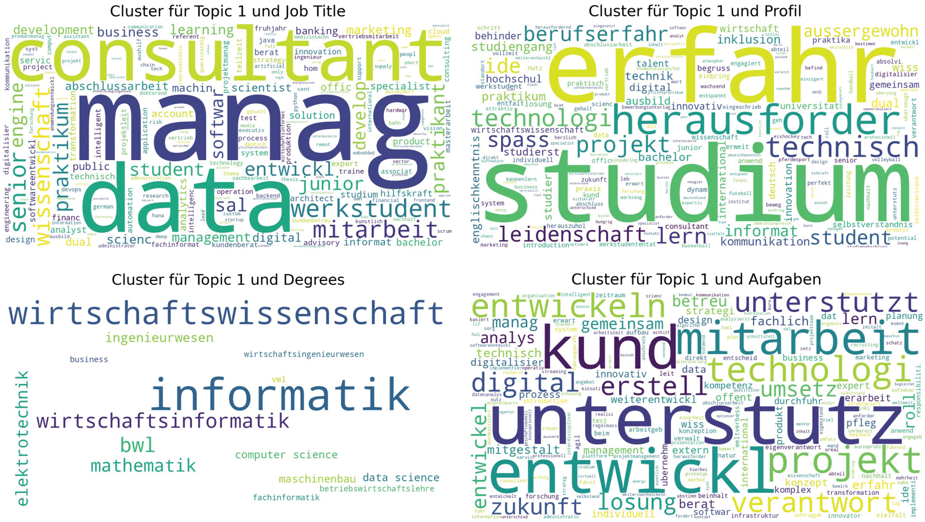


Figure 9.4: Most frequent keywords for persona 1.

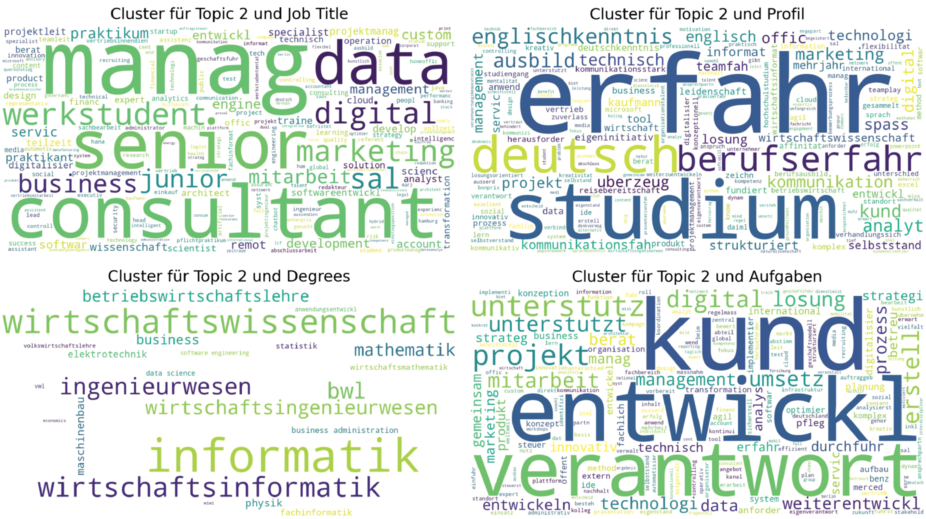


Figure 9.5: Most frequent key words for persona 2.

9.5.2.3 Persona 3 – Senior AI Architect

The third persona in the upper right corner of Figure 9.3 also takes responsibility for customers. But compared to persona 2, it focuses less on management tasks. The key terms associated with this persona highlight senior roles with a background in engineering and data science. The results concerning the profiles suggest that this persona has technical experience in data science and possesses competencies in technical domains, such as Python. In line with this, the degrees associated with this persona are primarily technical, emphasising information technology, computer science, mathematics and engineering. The tasks allocated to this persona suggest project-related responsibilities in the development of AI solutions, including consulting, again underlining the interpretation as a senior role with a strong technical focus. We give this persona the designation "Senior AI Architect." The word clouds in figure 9.6 provide an illustration of the most frequent keywords for persona 3.

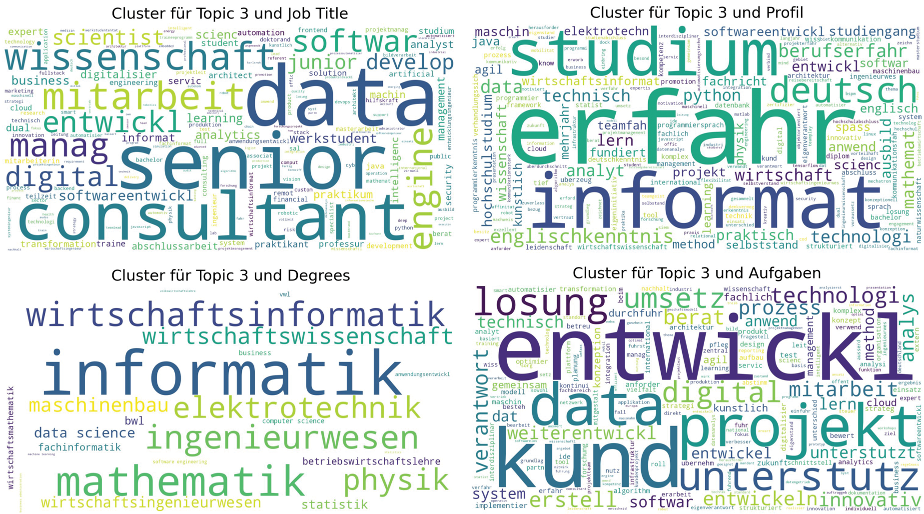


Figure 9.6: Most frequent keywords for persona 3.

9.5.2.4 Persona 4 – AI Developer

Persona 4 in the lower right corner of Figure 9.3 has a clear technical focus but is not involved in customer relationship management. The desired degrees are related to computer sciences and the tasks to be carried out are related to software development. A clear scientific background is necessary and desired experiences are related to specific programmes and frameworks. We describe this persona as "AI Developer." The word clouds in figure 9.7 provide an illustration of the most frequent keywords for persona 4.

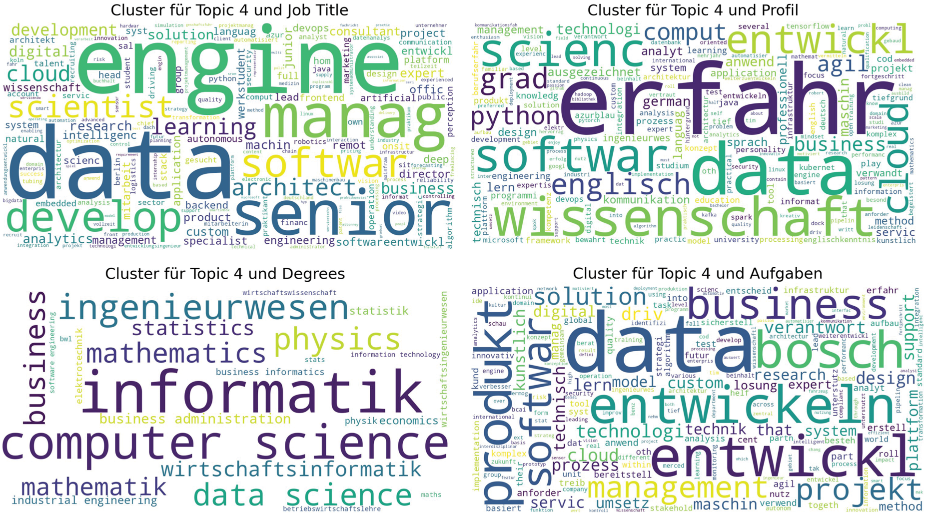


Figure 9.7: Most frequent keywords for persona 4.

9.5.3 Consolidation of the Results

The personas are developed from the interplay between desired profiles, job titles, tasks and further analyses such as an analysis of the context of the word "experience" [German: "Erfahrung"]. The consolidated results concerning the four personas are presented in Figure 9.8.

It seems that different personas are necessary to foster the diffusion of AI. The labour market, and in particular the growing field of AI-related jobs, hence provides opportunities for candidates with different backgrounds. Moreover, the personas do not necessarily have to be seen as separate individuals, but might also be interpreted as career-related development paths over time. Thus, a development from persona 1 to persona 2, 3 or 4 may occur as a person's career path is unfolding: Starting from the junior position, a person having studied business or economics a person can become a Senior AI Manager later, provided that the necessary technical or statistical knowledge or experience is available. However, the possible career paths in AI jobs seem to be more diverse for persons who have studied computer sciences or related topics. They might become a Senior AI architect, an AI developer or a Senior AI Manager. However, any interpretation of the career paths at this point is speculative and does not emerge from the data itself.

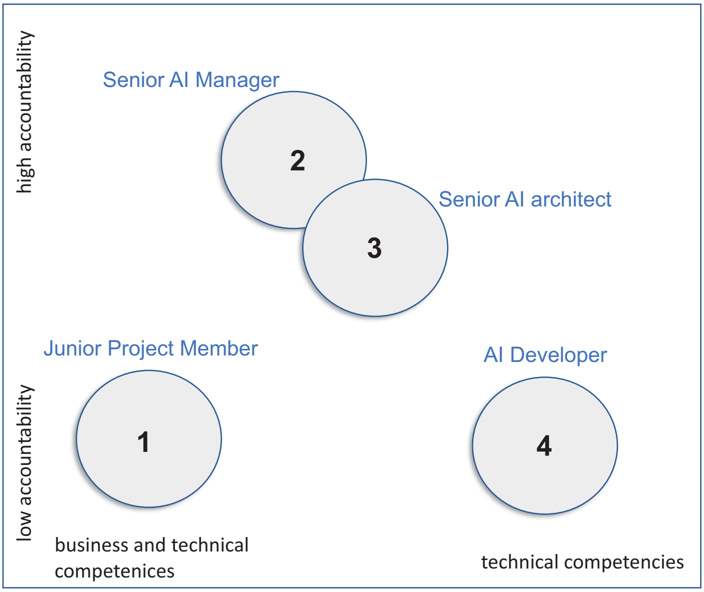


Figure 9.8: Interpretation of the LDA Model Solution for four personas.

9.5.4 Robustness

The LDA reacts sensibly to changes in the corpus. For a robustness check, we repeated the analysis ten times with 90% of the data used (i.e., 10% of the job postings were randomly removed in each of the ten runs), showing that the results of the LDA are to some extent arbitrary. As changes in individual words have a strong effect on the LDA algorithm, the four topic-solution presented before could not be replicated perfectly. In seven out of ten runs, the descriptions of a persona change slightly; in three out of ten runs, the generated topics do not match the personas described in this chapter. We expect that analyses with a larger database will generate more robust results.

9.6 Discussion

This study offers insights into the types of candidates and personas that are currently being recruited in the field of AI. This may not only add value for recruiting companies and current candidates but also help students in choosing their educational pathways. However, this study does not come without limitations. One limitation lies in the LDA and the assumption, that the words are independent from each other. This became obvious with the word “Erfahrungen” [English: Experience], where “Erste Erfahrungen” [English: first experiences] is related to a different persona than “langjährige Erfahrungen”

[English: long-term experiences]. Therefore, we checked the consistency with the persona descriptions manually. Another limitation lies in the adoption of a "naïve approach," as we assume that each topic also represents an ideal profile. Therefore, we also employed a DBSCAN cluster algorithm after the LDA, but this combination did not improve the results. Future work should encompass these shortcomings, using larger databases and word embeddings, as used in the Top2Vec algorithm (Angelov, 2020). Also, it might be expected that the new generative pre-trained transformers (GPT) models from OpenAI (Brown et al., 2020; OpenAI, 2023), Google (Vaswani et al., 2017) and others also lead to new approaches for topic modelling and classification. Also, for a comparison of countries with different languages over time, GPT embeddings could be employed.

In addition to the practical value of the results to companies, candidates, and researchers interested in the impact of AI on the labour market, this study also offers methodological insights into the usefulness of the concept of personas. A (candidate) persona is a hypothetical person who represents the desired candidate (Rippler, 2022). The concept was adopted to facilitate the presentation of the ideal-type candidates, as it reflects an intuitive and clustered interpretation of the candidates being sought. When used in development processes, personas are deliberately created by organisations to reflect the optimal candidates a company is looking for. The personas are hence created based on organisational requirements. In this study, the concept of a persona is used to retrospectively draw a comprehensive conclusion about ideal-type candidates based on the requirements identified in various job postings. The derived personas hence do not represent requirements of a specific organisation but at labour market level. The application of personas in this study has proven to be a useful construct for presenting the diverse results regarding ideal-type candidates in a concise and intelligible manner. The approach taken in this study is therefore considered to be promising and may be applicable in future research to identify profiles that are sought for in the labour market and, possibly, how these profiles change over time.

9.7 Summary

From an explorative analysis of online job postings for AI jobs in Germany, four personas could be derived. These personas are designated as Junior Project Member (1), Senior AI Manager (2), Senior AI Architect (3) and AI Developer (4). The personas are differentiated according to their accountability and the required competencies. The chosen designations are based on experts' interpretation and serve as a suggestion for generic labels. They do not necessarily correspond with the various job titles used by companies in their job postings in the underlying data set. Assuming that different personas might be necessary for a successful AI-related project, the results can be interpreted as a hint for a project-based diffusion of AI in companies.

Bibliography

Acemoglu, D., Autor, D. H., Hazell, J., & Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, *40*(S1), 293–340. https://doi.org/10.1086/718327.

Acemoglu, D., & Restrepo, P. (2022). A task-based approach to inequality. IFS Deaton Review of Inequalities, March 2022. https://ifs.org.uk/inequality/wp-content/uploads/2022/03/A-task-based-approach-to-inequality-IFS-Deaton-Review-of-Inequalities-1.pdf (accessed 9 July 2023).

Aghion, P., Antonin, C., Bunel, S., & Jaravel, X. (2022). The effects of automation on labor demand: A survey of the recent literature. CEPR Discussion Paper Series (DP16868). https://repec.cepr.org/repec/cpr/ceprdp/DP16868.pdf (accessed 9 July 2023).

Angelov, D. (2020). Top2Vec: Distributed representations of topics. https://arxiv.org/pdf/2008.09470.pdf (accessed 9 July 2023).

Anton, E., Behne, A., & Teuteberg, F. (2020). The humans behind Artificial Intelligence – An operationalisation of AI competencies. ECIS 2020 Proceedings. Research Papers. https://aisel.aisnet.org/ecis2020_rp/141 (accessed 9 July 2023).

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, *118*(4), 1279–1333. https://doi.org/10.1162/003355303322552801.

Blei, D. M. (2012). Probabilistic topic models. *Communication of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Bouma, G. (2009). Normalized (Pointwise) Mutual information in collocation extraction. Proceedings of Gesellschaft Für Sprachtechnologie Und Computerlinguistik (GSCL), 30, 31–40. https://svn.spraakdata.gu.se/repos/gerlof/pub/www/Docs/npmi-pfd.pdf (accessed 9 July 2023).

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, S., & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. https://doi.org/10.48550/arXiv.2005.14165.

Card, D. E., & DiNardo, J. E. (2002). Skill-biased technological change and rising wage inequality: Some problems and puzzles. *Journal of Labor Economics*, *20*(4), 733–783. https://doi.org/10.1086/342055.

Cooper, A. (1999). The inmates are running the asylum. In: U. Arend, E. Eberleh, & K. Pitschke (Eds.), *Software-Ergonomie '99: Design von Informationswelten*. Vieweg & Teubner Verlag. https://doi.org/10.1007/978-3-322-99786-9_1.

Debortoli, S., Müller, O., & vom Brocke, J. (2014). Comparing business intelligence and big data skills: A text mining study using job advertisements. *Business & Information Systems Engineering*, *6*(5), 289–300. https://doi.org/10.1007/s12599-014-0344-2.

De Mauro, A., Greco, M., Grimaldi, M., & Ritala, P. (2018). Human resources for big data professions: A systematic classification of job roles and required skill sets. *Information Processing and Management*, *54*(5), 807–817. https://doi.org/10.1016/j.ipm.2017.05.004.

Gurcan, F., & Cagiltay, N. E. (2019). Big data software engineering: Analysis of knowledge domains and skill sets using LDA-based topic modeling. *IEEE Access*, *7*, https://doi.org/10.1109/ACCESS.2019.2924075.

Hoffman, M. D., Blei, D. M., & Bach, F. (2010). Online learning for Latent Dirichlet Allocation. *Advances in Neural Information Processing Systems*, *23*, 856–864. https://papers.neurips.cc/paper/2010/file/71f6278d140af599e06ad9bf1ba03cb0-Paper.pdf (accessed 9 July 2023).

Hornstein, A., Krusell, P., & Violante, G. L. (2005). The effects of technical change on labor market inequalities. In: P. Aghion & S. Durlauf (Eds.), *Handbook of Economic Growth* (Vol. 1B, 1275–1370). Elsevier.

Koong, K. S., Lui, L. C., & Lui, X. (2002). A study of the demand for information technology professionals in selected internet job portals. *Journal of Information Systems Education*, *13*(1), 21–28. http://jise.org/Volume13/n1/JISEv13n1p21.pdf.

Lehnert, K., Goupil, S., & Brand, P. (2021). Content and the customer: Inbound ad strategies gain traction. *Journal of Business Strategy*, *42*(1), 3–12. https://doi.org/10.1108/JBS-12-2019-0243.

Litecky, C., Aken, A., Ahmad, A., & Nelson, H. J. (2010). Mining for computing jobs. *IEEE Software*, *27*(1), 78–85. https://doi.org/10.1109/MS.2009.150.

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing Semantic Coherence in Topic Models. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, 262–272, Edinburgh, Scotland, UK. Association for Computational Linguistics.

Newman, D., Han Lau, J., Grieser, K., & Baldwin, T. (2010). Automatic Evaluation of Topic Coherence. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 100–108, Los Angeles, California. Association for Computational Linguistics.

Nielsen, L. (2019). *Personas – User Focused Design*. Springer. https://doi.org/10.1007/978-1-4471-7427-1

OpenAI. (2023). GPT-4 technical report. https://doi.org/10.48550/arXiv.2303.08774 (accessed 9 July 2023).

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. https://doi.org/10.48550/arXiv.1908.10084 (accessed 9 July 2023).

Rippler, S. (2022). *Das Persona-Prinzip. Erfolgreiches Recruiting mit Candidate Personas*. Springer Gabler. https://doi.org/10.1007/978-3-658-38979-6.

Röder, M., & Both, A. (2015). Exploring the space of topic coherence measures. Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, 399–408. http://doi.org/10.1145/2684822.2685324

Samek, L., Squicciarini, M., & Cammeraat, E. (2021). The human capital behind AI: Jobs and skills demand from online job postings. *OECD Science, Technology and Industry Policy Papers*, 120. https://doi.org/10.1787/2e278150-en.

Squicciarini, M., & Nachtigall, H. (2021). Demand for AI skills in jobs: Evidence from online job postings. *OECD Science, Technology and Industry Working Papers*, 2021/03. https://doi.org/10.1787/3ed32d94-en.

Ternikov, A. (2022). Soft and hard skills identification: insights from IT job advertisements in the CIS region. *PeerJ Computer Sciences*, 8. http://doi.org/10.7717/peerj-cs.946.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30. https://doi.org/10.48550/arXiv.1706.03762.

Verma, A., Lamsai, K., & Verma, P. (2022). An investigation of skill requirements in artificial intelligence and machine learning job advertisements. *Industry and Higher Education*, *36*(1), 63–73. https://doi.org/10.1177/0950422221990990.

Ed Dandalt

Chapter 10
Artificial Intelligence and Care Leaders: A Critical Perspective

Abstract: This essay addresses the question of why artificial intelligence (AI) is less likely to make the roles of physicians as care leaders obsolete. The argument focuses on the uniqueness and relevancy of the roles of physicians as managers and clinicians. This argument is supplemented by a discussion on the pushover factor that the institutional environment could play in maintaining the role privilege of physicians vis a vis the rollout of AI in healthcare organisations. The potential outcomes represented by AI are also discussed.

Keywords: medical AI, AI in healthcare

10.1 Management and Leadership Roles

AI is less likely to make the roles of physicians as care leaders obsolete because it cannot automate the dualistic component of their workflow system. Their role set is a blending of clinical and management skills. Clinical skills are particularly the ones that are the most difficult to make obsolete (Claiborne, 2018). Although AI will not erode the job security of physicians in the future, I still associate this technological change with beneficial outcomes. AI is changing the knowledge management systems used by physicians. They are increasingly relying more on virtual data sources rather than textbooks to aid in diagnosis and treatment. As such, AI represents an asset in reducing the cognitive labour of physicians as it is de-emphasising memorisation in the medical profession. Because of that, experts within the medical field posit that the way physicians work and are trained needs to be reimagined (Claiborne, 2018).

Plus, AI will free physicians from administrative coordination and control tasks as organisational research suggests that management work steals time from clinical work among physicians. This is a possible outcome because it is predicted that AI will sooner be able to do the administrative tasks that consume much of managers' time, faster and better (Radin et al., 2020). Arguably, AI could soften the bureaucratisation or restratification of physicians. AI could democratise knowledge sharing and the flow of information between health professionals. This democratisation could take

Ed Dandalt, Faculty of Human and Social Sciences, Wilfried Laurier University, Waterloo, Canada, e-mail: e.dandalt@wlu.ca

https://doi.org/10.1515/9783110790320-010

the form of providing intelligent systems that will enable work processes and decisions that are not disjointed or fragmented.

But at the same time, it is possible that the automation of certain management tasks by AI such as electronic performance monitoring (EPM) could reduce the periphery of physicians' control over the management of care and their staff members. There are some indicators suggesting that healthcare organisations are moving in that direction, considering that computers can now record precisely the work that healthcare workers (e.g., nurses) perform in hospitals. But this organisational change is far from leading AI to adequately perform the core leadership roles of physicians such as empathy and active listening. From a theoretical stance, this rationale is supported by the work of scholars such as Montemayor, Halpern & Fairweather (2022) which suggests that AI cannot match the level of empathy that physicians embody during therapeutic care. The art of caring will be much more difficult to replace with technology (Claiborne, 2018).

10.2 Market and Social Forces

The present attitude of patients towards AI suggests that the roles of physicians as care leaders will not become obsolete as a recent study completed by researchers at the Pew Research Center (Tyson et al., 2023) shows that 60% of patients would not be comfortable with clinicians relying on AI in the provision of care services. There are 33% of those who believe it would lead to worse outcomes and 27% think it would not make much difference. Only 38% of patients believe that AI would lead to better health outcomes. Plus, the World Health Organization (WHO) posits that the discourse of technological solutionism espoused by the proponents of AI presents a risk for public health. By that it means that the benefits associated with AI for improving public health are overestimated. One of the arguments raised by WHO (2021) to explain this caution is that the unregulated use of AI could subordinate the rights of patients to the powerful commercial interests of technology firms.

The above rationale can be explained from the stance of Feenberg's critical theory of technology (2005). This theory is concerned with the threat to human agency posed by the technocratic system that dominates modern societies. Leveraging on this theory, I argue that AI can only pose a serious risk to physicians once patients start believing that the roles of physicians are irrelevant in an AI-powered healthcare system. In Bourdieusian terms, there needs to be a paradigm shift in the collective habitus vis a vis of the symbolic power represented by physicians. In support of this argument, a recent study led by researchers from the University of Arizona Health Sciences found that 52% of patients prefer physicians rather than AI for treatment. Findings from studies by researchers (Longoni, Bonezzi, & Morewedge, 2019) at the New York Stern School Business, and by Promberger and Baron (2006) show that patients trust physicians more than AI. These patients' resistance to AI-powered care processes derives from their perceptions of "uniqueness neglect" or the belief that AI cannot effectively address the intricacies of their individualised care needs in the same way that physicians can. At the core of this receptivity differential is the belief that following physicians' recommendations reduced their feeling of responsibility more than following those generated by AI. Arguably, the adoption of medical AI partly depends on consumer receptivity in society at large.

Adding to the above argument, the ongoing orientations of government policies suggest that the expertise of physicians is in higher demand, especially in primary care. In the USA, for example, it is projected that by 2034, demand for physicians will exceed supply by a range of between 37,800 and 124,000 physicians. One of the driving forces of this labour shortage is that the US population will increase by 10.6%, namely from 328 million to 363 million (AAMC, 2020). The population aged 65 years old and over will grow by 42.4% and will need to have access to more clinical services such as palliative care. Rather than relying on AI, policymakers in the United Kingdom (UK) are trying to address the problem of medical labour shortage by allowing the recruitment of foreign physicians. In 2022, 46% of physicians who joined the National Health Service (NHS), respectively 12,148, were trained abroad or outside of the European Union. Germany faces the same issue as it is predicted that there is a need to add 106,000 physicians by 2030 (Klingler & Marckmann, 2016). This labour supply crisis has led German hospitals to recruit foreign physicians with the help of outsourcing agencies to help fill this gap. In Taylorian words, these indicators insinuate that the efficiency of AI might not be perceived by policymakers as the 'one best way' of addressing labour shortage in the healthcare industry.

Furthermore, the World Economic Forum posits that by 2025, AI will contribute to the creation of at least 12 million more jobs across most industries than they eliminate (Moret, 2023). This means that, contrary to the proponents of the job displacement theory, AI could add more jobs in the medical profession. All in all, my argument aligns with the theory of automatic reabsorption of technologically displaced labour (Acemoglu & Restrepo, 2019). This economic theory argues that innovations are not the cause of unemployment. This view is also supported and shared by contemporary economists such as Acemoglu and Restrepo (2019) who stated that "The history of technology is not only about the displacement of human labour by automation technologies. Instead, the displacement effect of automation has been counterbalanced by technologies that create new tasks in which labour has a comparative advantage. The reinstatement effect is the polar opposite of the displacement effect and directly increases the labour share as well as labour demand" (Acemoglu & Restrepo, 2019, p. 4).

10.3 Potential Outcomes and “McDonaldisation”

There is a possibility that AI could lead to the de-personalisation of physician–patient relationships. This argument is supported by the rationale of Ben Seligman (1965) who argued that technology is limited because, under the control of automation, work becomes a mechanical reaction, loses its spontaneity, creativeness and is converted into automatic behaviours. Meanings disappear as work takes on the character of continuous improvement. When it does, those humans utilised by the process become mere automatons. This phenomenon is already occurring in relation to the “McDonaldisation” of medicine as Dorsey and Ritzer (2016) argue that electronic medical control systems are dehumanising medical practices with physicians spending 40% of their working time on computers and only 12% of their time with clients.

McDonaldisation refers to the introduction of the business principles (efficiency, calculability, predictability and control) of fast-food industries in clinical work processes. In healthcare management literature, it is defined as the control of healthcare professionals by nonhuman technology and standardisation. As such, an AI-engineered “McDonaldisation” could contribute to the de-professionalisation of physicians, considering that organisational research suggests that “McDonaldisation” leads to job deskilling and workforce casualisation.

Moreover, AI could be beneficial to governments in advanced economies by serving as a strategic tool to reduce budget deficits in the healthcare sector as labour scholars (Dau-Schmidt, 2015; Estlund, 2018) argue that technological change is more beneficial to employers than to workers. It saves labour costs, facilitates job outsourcing and lowers employment standards at the expense of workers. As a result, workers have seen their fringe benefits, compensation and job security being curbed. This economic reality is mostly experienced by workers with low skills and low educational achievements. But workers with higher skills such as physicians may no longer be immune from loss in collective bargaining rights once AI becomes better equipped at performing some of their tasks. This hypothetical outcome could occur in the form of wage loss in the fee-for-service (FFS) system as physicians will no longer be able to charge for some of the clinical tasks performed by AI. This trickle-down effect is already occurring in relation to telehealth. A paper by Wilson et al. (2017) pointed out that compared to face-to-face services, the average reimbursements for telehealth services are significantly lower for physicians. Telehealth reimbursements for evaluation and management of clients with low complexity are 30% lower than the corresponding non-telehealth services.

In countries such as France with a unionised medical workforce, any attempts from government leaders to overhaul the employment rights of physicians through AI-powered “McDonaldisation” could be met with union resistance. Medical associations that bargain for the self-interests of their rank-file members will make it difficult for governments to achieve such a change agenda, considering that systems of governance in hospitals are bicameral (power shared between physician leaders and non-clinical executives). This inference is substantiated by the discourse held by Charles Killingsworth and Terry Moe on the political economy of organised labour. Killingsworth (1962) noted that full automation is adopted most extensively in industries where collective bargaining is rare. Moe (2011) argues that unions always play "the politics of blocking" to protect their rank-file members against threatening reforms or changes. This potential outcome is plausible and can be argued from a pluralist perspective of employment relations. This theory is the most popular model used in organisations and suggests that management–employee conflicts are unavoidable and inherent to employment relations. Conflicts are the outcome of competing interests between different groups and imperfect labour markets (Budd et al., 2004).

Furthermore, research suggests that the adoption process of technology in modern organisations is not linear or unilateral because it involves a wide range of actors and interests, which operate under different incentives (Gelijns & Rosenberg, 1994). This means that at the individual level, physicians will only support AI to the level that it does not pose a threat to their job security. This is a plausible outcome as this democratic model of technology adoption gives users the power to propose incremental changes to developers through feedback mechanisms. Arguably, the voice of physicians as stakeholders will always be part of innovation processes in healthcare organisations (Waring & Bishop, 2013). In conclusion, my above argument can be reduced to the simple idea that the displacement of physicians by AI is science fiction.

Bibliography

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labour. *Journal of Economic Perspectives*, *33*(2), 3–30.

Association of American Medical Colleges (AAMC). (2020). The complexities of physician supply and demand: Projections from 2018 to 2033. Washington, DC: AAMC.

Budd, J., Gomez, R., & Meltz, N. (2004). Why a balance is best: The pluralist industrial Relations paradigm of balancing competing interests. In: Kaufman, B.E. (Ed.), *Theoretical Perspectives on Work and the Employment Relationship* (pp. 195–227), Champaign, IL: Industrial Relations Research Association.

Claiborne, J. (2018). Anticipating and training the physician of the future: The importance of caring in an age of artificial intelligence. *Academic Medicine*, *93*(8), 1105–1106.

Dau-Schmidt, K. (2015). The implication of new information technology on the employment relationship and the relevance of the NRLA. *Emory Law Journal*, *64*,1583–1609.

Dorsey, R.M., & Ritzer, G. (2016). The McDonaldisation of medicine. *JAMA Neurology*, *73*(1), 15–16.

Estlund, C. (2018). What should we do after work? Automation and employment law. *Yale Law Journal*,128, 245–326.

Feenberg, A. (2005). Critical theory of technology: An overview. *Tailoring Biotechnologies*, *1*(1), 47–64.

Gelijns, A., & Rosenberg, N. (1994). The dynamics of technological change in medicine. *Health Affairs*, *13*(3), 28–46.

Killingsworth, C. (1962). Industrial relations and automation. *The Annals of the American Academy of Political and Social Science*, *340*(1), 69–80.

Klingler, C., & Marckmann, G. (2016). Difficulties experienced by migrant physicians working in German hospitals: A qualitative interview study. *Human Resources for Health*, *14*(57), 1–13.

Longoni, C., Bonezzi, A., & Morewedge, C. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, *46*(4), 629–650.

Moe, T. (2011). *Special Interest: Teachers Unions and America's Public Schools*. Washington, DC: Brookings Institution Press.

Montemayor, C., Halpern, J., & Fairweather, A. (2022). In principle obstacles for empathic AI: Why we can't replace human empathy in healthcare. *AI & Society*, *37*, 1553–1559.

Moret, B. (2023). Can automation pull us through the global labour shortage. Retrieved on November 8, 2023, https://www.weforum.org/agenda/2023/01/how-automation-will-pull-us-through-the-labour-shortage-davos23/.

Promberger, M., & Baron, J. (2006). Do patients trust computers? *Journal of Behavioral Decision Making*, *19*(5), 455–468.

Radin, J., Gordon, R., Elsner, N., & Mukherjee, D. (2020). Rethinking the physician of the future: Embracing new technologies, empathy, and new models of care. Deloitte, pp. 1–16.

Seligman, B. (1965). On work, alienation, and leisure. *American Journal of Economics & Sociology*, *24*(4), 337–360.

Tyson, A., Pasquini, G., Spencer, A., & Funk, C. (2023). 60% of Americans would be uncomfortable with provider relying on AI in their own health care. Retrieved on November 8, 2023, https://www.pewresearch.org/science/2023/02/22/60-of-americans.

Waring, J., & Bishop, S. (2013). McDonaldisation or commercial re-stratification: Corporatisation and the multimodal organisation of English doctors. *Social Science & Medicine*, *82*, 147–155.

Wilson, F., Rampa, S., Trout, K., & Stimpson, J. (2017). Reimbursements for telehealth services are likely to be lower than non-telehealth services in the United States. *Journal of Telemedicine & Telecare*, *23*(4), 497–500.

World Health Organization (WHO). (2021). *Ethics and Governance of Artificial Intelligence for Health*. WHO: Geneva.

Part 6: **Artificial Intelligence and Humans**

Zunera Rana, Jessica Roemer, Thomas Pitz, Joern Sickmann

Chapter 11 Public Perception of Artificial Intelligence: A Systematic Evaluation of Newspaper Articles Using Sentiment Analysis

Abstract: In our research, we focus on the reporting on artificial intelligence (AI) and related terms in newspaper articles and determine if the sentiment tone of the articles is positive, negative or neutral. We hypothesise that the tone of reporting about AI can shape consumer opinions and usage patterns since it increases familiarity and can therefore indirectly impact the level of trust towards AI among German consumers. We use a quantitative approach consisting of sentiment analysis of the leading newspapers (in terms of number of subscribers) in Germany to test our hypothesis and conclude that the newspapers have mostly published positive or neutral articles albeit the frequency of articles published on the theme of AI has decreased over the past 2 years. However, only a small percentage of articles discuss issues related to consumers and consumer protection, which might be directly relevant to the readers of the newspapers.

Keywords: artificial intelligence, algorithms, sentiment analysis, newspapers, public opinion

11.1 Introduction

Artificial intelligence (AI) and algorithms have long been a topic of interest for the public and have been portrayed in movies (e.g., *Terminator* and *Star Wars*) and media quite often as well. Over time, the use of AI in our everyday lives has become

Funding: The project is supported (was supported) by funds of the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support programme.

Zunera Rana, Radboud University, Faculty of Social Sciences, e-mail: zuneraahmad.rana@ru.nl
Jessica Roemer, Faculty of Society and Economics, Rhine-Waal University of Applied Sciences, e-mail: jessica.roemer@hochschule-rhein-waal.de
Thomas Pitz, Faculty of Society and Economics, Rhine-Waal University of Applied Sciences, e-mail: thomas.pitz@hochschule-rhein-waal.de
Joern Sickmann, Faculty of Society and Economics, Rhine-Waal University of Applied Sciences, e-mail: joern.sickmann@hochschule-rhein-waal.de

https://doi.org/10.1515/9783110790320-011

more and more prominent. From health care to human resource systems to financial advisory and driverless cars, algorithms are slowly sneaking into our lives. Despite the convenience and (in some cases) accuracy that AI brings with it, there is still a lot of variation in the degree of individual adherence to algorithmic advice; scepticism and algorithmic aversion can be observed among groups of the public about it (Dietvorst, Simmons & Massey, 2015). There are several reasons behind this scepticism, including issues like data privacy (Fast & Jago, 2020), efficacy (Fenneman et al., 2021), ethical considerations (Lee, 2018), fairness (Castelo, Bos & Lehmann, 2019) and transparency (Dzindolet et al., 2001) of algorithms.

Newspaper articles and media can play an important role in shaping public opinion particularly about new technologies. The role of media in creating awareness and engagement towards scientific advancement has long been a topic of discussion in the literature (Nelkin, 1987; 1998) with the scientific community mostly holding a negative stance towards the criticism posed by media (Garvey, 2018). Similarly, within the AI community, there is a strong belief that the negative media coverage of AI and the use of imagery from the movie *Terminator* has played a significant role in increasing the scepticism and public concern towards algorithms and AI (Garvey & Maskal, 2019). This has in turn hindered public engagement on AI by creating a psychocultural barrier (Dotson, 2015; Garvey, 2018).

This negative view of the role of media has been somewhat explored in the literature by reviewing articles from international newspapers, including mainly *The New York Times* (Fast & Horvitz, 2016; Garvey & Maskal, 2019) and from social media site and blogs (mainly Twitter). The results in the literature are mixed with most studies not finding clear evidence of the link between negative hype of AI and the low public engagement of it (see Section 12.2).

We have only been able to find one (non-peer-reviewed) study and one blog article on the sentiment in German newspapers on the topic of AI.[1] This is especially interesting because the German public relies heavily on newspapers to get information on current events. More than 11.72 million copies of traditional newspapers circulate every day in Germany (Bundesverband Digitalpublisher und Zeitungsverleger, 2022). Newspapers in Germany potentially play an important role in forming public opinion about important topics, including AI and algorithmic usage in everyday life. Through this chapter, we intend to add to this limited literature and explore the role of German newspapers in depicting positive or negative sentiments towards AI.

Our study is guided by four main research questions:

– What is the frequency of articles published on the theme of AI and algorithms?
– What has the sentiment tone of newspapers been about the theme of AI and algorithms?

1 Ozgun and Broekel (2021) have discussed innovation and technological news in German media but have not specifically focused on artificial intelligence.

- To what extent do AI and algorithm-related articles discuss consumers or consumer-related issues? Does the sentiment score differ for this subsample?
- Are the sentiments expressed by newspaper articles reflected in the public opinion of AI and algorithms?

To answer our research questions and have a more in-depth analysis, we limited our sample size to three German newspapers, each with a different orientation (tabloid, national and regional quality newspaper) targeting a different stratum of German population. After retrieving the articles from the sample newspapers, we opted for a sentiment analysis approach to determine if the articles portrayed the use of AI and algorithms in a positive, negative or neutral light. In addition to this, we also conducted robustness checks on our findings by comparing our results with regional and national German public opinion polls on AI and algorithms. We discuss our methodological approach in more detail in Section 12.3.

Through this research, we hope to add to the current literature on two levels: first, we want to explore the current sentiment in the German media about AI and algorithms, which has been almost missing from literature so far. Given that the EU hopes to be at the forefront of technological progress, our research can play a pivotal role in understanding the current sentiment of discussions in the media about technological changes related to algorithms and AI. Secondly, we hope to add to the bigger debate on the role of media in increasing or decreasing public engagement towards AI and algorithms.

We have divided the chapter into five sections: in Section 12.2, we explore the literature related to AI, algorithmic aversion and the role of media. This is followed by Section 12.3, which explains the methodology and results of the chapter. In Section 12.4, we analyse the results and compare them with other studies and public polls held in Germany and discuss broader conclusions from our results.

11.2 Sentiment Analysis and AI: A Literature Review

Over the last years, several problems and scandals with regard to AI have been published in the German press. Examples for negative press reports about AI are the open letter of warning about AI risks which was signed by more than 100 AI experts, such as Stephen Hawking and Elon Musk, and was addressed to the United Nations (Armbruster, 2017), the Facebook–Cambridge Analytica data scandal (Zeit Online, 2018), discrimination against women in Amazon's AI application system (Sackmann, 2018) and crashes of autonomous cars (Handelsblatt, 2021).

However, studies on the sentiment of news articles about AI in the USA report a positive sentiment towards AI (Fast & Horvitz, 2016; Garvey & Maskal, 2019). Nevertheless, while the results of a content analysis of five major American newspapers also support these findings by reporting that the share of articles discussing the benefits of AI is higher than the share of articles discussing the risks, they also report that the discussions of the disadvantages and risks are more specific (Chuan, Tsai & Cho, 2019). Moreover, while the general sentiment in AI may be positive, dividing the news articles on AI into sub-topics leads to a more precise analysis of the sentiment and reveals differences between the sentiments of the articles of the various subtopics. For example, while the general sentiment of AI articles was reported to be positive by several studies (Fast & Horvitz, 2017; Garvey & Maskal, 2019), articles on the ethics of AI were found to have a neutral sentiment (Ouchchy, Coin & Dubljevic, 2020).

In the recent years, the frequency of press articles on AI has decreased. Garvey and Maskal (2019) report a decrease of news articles on AI since 2016. This is somewhat surprising as the number of scientific publications in English on AI increased from 194,194 publications in 2016 to 334,497 publications in 2021 (Zhang et al., 2022). This indicates different trends for scientific publications and non-scientific publications, which might reflect a declining interest of the public in the topic. Garvey and Maskal (2019) offer another possible explanation for this difference by stating that public engagement is restricted by measures as banning journalists from workshops and tutorials on AI conferences (Shead, 2018).

For the German media market, similar peer-reviewed studies do not exist to the best of our knowledge. We are only aware about one study of the Bertelsmann Stiftung (Fischer & Puschmann, 2021), one blog article including a sentiment analysis for German newspapers on AI (Illner, 2022) and the AI Monitor 2022, where the tonality of articles is part of the AI index (Bundesverband Digitale Wirtschaft e.V., 2022). In his blog article, Illner comes to a similar result as Garvey and Maskal (2019) by analysing the press articles of 3000 publications and reports a decrease in press articles on AI from January 2020 to January 2022. Fischer and Puschmann (2021) come to a different result and report that the number of articles in the leading German newspapers continued to increase until 2020. A decreasing frequency of newspaper articles on AI may seem surprising, considering the fact that AI is being used more and more. From 2018 to 2022, the percentage of companies that use AI increased from 11% to 37% (Bitkom Research, 2022).

According to the non-peer-reviewed blog article from Illner (2022), the share of negative press articles is lower than the share of positive press articles on AI and even decreased from 20.5% in 2020 to 17% in 2021. Simultaneously, the share of positive press articles increased from 32.3% in 2020 to 35% in 2021 (Illner, 2022). Furthermore, Fischer and Puschmann (2021) report that more articles and posts from leading German newspapers, subject blogs and Twitter messages are positive than neutral or negative towards AI. However, the Bundesverband Digitale Wirtschaft e.V. (2022) reports that the proportion of negative and positive articles from German newspapers between 2019 and 2022 was balanced but that there is a small increase in positive articles.

The German population's attitude towards AI has changed over time; more precisely, it has become more positive since 2019 according to a survey among the German population (TÜV-Verband, 2021). While in 2019, 28% of the respondents had a negative attitude towards AI, only 14% of the respondents in 2021 had a negative attitude towards AI (TÜV-Verband, 2021). The share of people with a positive attitude towards AI increased from 46% in 2019 to 51% in 2021 (TÜV-Verband, 2021).

Several studies have reported that the sentiment of media influences the public opinion about AI (Fast & Horvitz, 2016; Nader et al., 2022; Ouchchy, Coin & Dubljevic, 2020). The sentiment of the media is particularly important for shaping public opinion on new, unfamiliar technologies, such as AI, where many people feel unsure about the risks and benefits (Chuan, Tsai & Cho, 2019). It can thus be assumed that more positive press reports about AI would promote a more positive public opinion about AI. Furthermore, in a survey conducted by TÜV-Verband in 2019 in Germany, 59% of the respondents stated that positive reporting on AI in the media would increase their trust in products and applications using AI (TÜV-Verband, 2021).

The sentiment analysis is considered as one of the most accurate methods to evaluate text in terms of its sentiment and emotions (Hossain et al., 2021). However, using this method to analyse texts on AI has some limitations. The term artificial intelligence has no universal, specific definition and people have different understandings of what AI is. As a result, some people associate certain terms with AI and others do not associate those terms with AI. Therefore, AI is a difficult topic to conduct a sentiment analysis on and the search words must be considered carefully (Fast & Horvitz, 2016). Additionally, the result of the sentiment analysis can be distorted by rhetorical stylistic techniques such as irony, ambiguous words, abbreviations or typing errors (Hossain et al., 2021). For this reason, some studies decided to have people additionally evaluate the sentiment of the articles (Fast & Horvitz, 2016).

11.3 Methodology and Results

As mentioned in Section 12.1, we opted for a sentiment analysis approach as our main methodology. However, before we discuss the sentiment analysis, we explain our sample and data collection techniques in the following sub-sections.

11.3.1 Data Sampling and Collection

To have a more comprehensive overview of the news in Germany, we opted for three newspapers: Bild, Sueddeutsche Zeitung (SZ) and Rheinische Post (RP). Bild and SZ are the top two most read newspapers in Germany in terms of number of subscribers with Bild having 5.51 million daily unique users (Bild, 2022) and SZ having 1.33 million subscribers in 2022 (Axel Springer, 2022). Given the reach of non-academic publications, it is advisable to include tabloids and soft news in a sentiment analysis (Vicsek, 2011). Soft news shows the highest engagement rate and thus has a greater reach compared to scientific papers and hard news (Vicsek, 2011). For this reason, we have selected SZ as a newspaper with rather hard news and the tabloid newspaper Bild reports on soft news more often than hard news. At the same time, we wanted to analyse a regional newspaper as well to compare and contrast the differences between the articles published on regional and national level. For this reason, we opted for RP – since it is one of the biggest local newspapers in North Rhein-Westphalia, the most populous province of Germany.

We analyse the time period from 2010 to 2022 and select this time frame for reasons of data availability. For our initial analysis, we searched for the term "artificial intelligence" or "Künstliche Intelligenz" in either the title or main text of the article. The search was then extended to a variation of words related to AI and algorithms, which can be seen in Table 11.1. We used a web scraping tool from R to access the articles from each website after getting paid subscription for the newspapers. The web scraping tool allowed us to download article titles, text and respective dates on which the article was posted.

Table 11.1: List of keywords for AI-related article search (Fischer & Puschmann, 2021).

AI and algorithm-related keywords
ADM
Algorithmus
Algorithmen
Algorithmische Entscheidungen
Artificial Intelligence
Automatisierte Entscheidungsfindung
Künstliche Intelligenz
KI
Maschinelles Lernen
Machine learning
Maschinenlernen
Maschinenlernverfahren

To ensure that only articles that address the topic of artificial intelligence were captured, we tried different variations of the keyword artificial intelligence. For example, we checked for "artificial intelligence", "artificial AND intelligence", "artificial OR intelligence" etc. The "AND" and "OR" research functions allowed us to separate articles that were captured by the web scraping tool but only contained one of the two words from the keyword "artificial intelligence". This means that we were able to omit articles that discussed only "artificial" or "intelligence".

A total of 11,319 articles from all three newspapers combined were identified from the selected keywords and downloaded. The downloaded articles were checked for duplicates (based on the title of the news); as a result, 9487 articles remained in the final dataset. We also observed that in some cases, AI or algorithms were not exactly the theme of the article, even though the keyword was mentioned in the text. For example, in one instance, the article was identified as AI-related because it included a sentence "*The minister president said that Germany needs to invest more in artificial intelligence*" while the rest of the article was about the political policy of the said minister president, not focusing on AI at all. This led us to develop additional filter to ensure that the articles we are including in our analysis were indeed about AI.

As a first filter, we counted the number of times each keyword was appearing in each article. If it was only appearing once, the article was removed from the sample. If the keyword or a combination of keywords were appearing more than once, the articles were kept. We then did a second robustness check by manually selecting (randomly) and reading articles from our sample to ensure that they were indeed about the theme of AI or algorithms. A total of 45 articles (15 from each newspaper) were tested randomly. As a result of both checks, our sample size reduced significantly (from 9487 observations to 2240 observations). We discuss the data for each newspaper individually in more detail in the next section. Figure 11.1 gives a summary of the process.

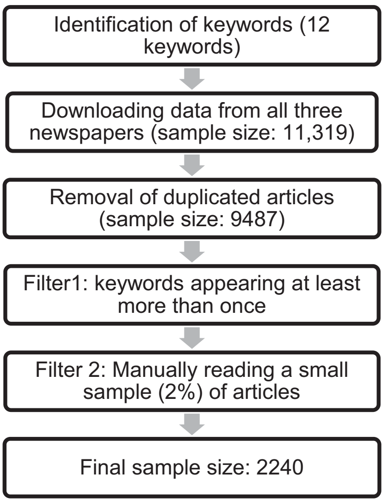


Figure 11.1: Data collection process. (compiled by authors.)

11.3.2 Summary of Data

Our sample size indicates that the highest number of articles published on the theme of AI and algorithms were published in the SZ (1589 articles), followed by the RP (351 articles). Bild published the lowest number of articles (300) between 2013 and 2022, as can be seen from Figure 11.2.

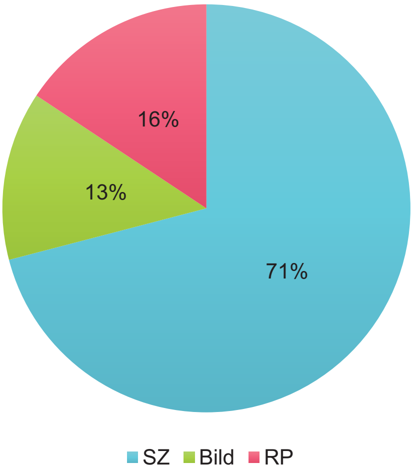


Figure 11.2: Share of articles on AI and algorithms between Bild, SZ and RP between 2013 and 2020.

Contrary to the conclusions drawn by Garvey and Maskal (2019), our data indicates an increase in newspaper articles from 2013 onwards, which peaked in 2019 (see Figure 11.3). Since 2019, the interest of newspapers in AI-related topics seems to have decreased relatively. One reason behind this could be the increasing ubiquity of AI; the term is taking a back seat in favour of terms related to concrete applications for example Chat-GPT3, DALL-E or other similar terms. The highest number of articles published were in 2018 and 2019, which could be attributed to the scandals related to Amazon and Facebook–Cambridge Analytica that became the talk of the town around that time. It is also interesting to see that in the case of Bild, the publication of AI-related articles remained steady between 2016 and 2019 while it increased drastically for SZ and increased only somewhat for RP.

When looking at the frequency of articles published on monthly basis, it was very interesting to observe that for all three newspapers, the average number of articles decreased in the month of August while the peeks differ significantly across the newspapers (for SZ and Bild, the highest number of articles published were in the months between October and December while for RP it was between May and June).

11.3.3 Sentiment Analysis

To answer our second research question, we opted for a sentiment analysis approach by making use of the syuzhet package (Jocker, 2020) in R, which makes use of four different dictionaries along with natural language processing techniques to determine

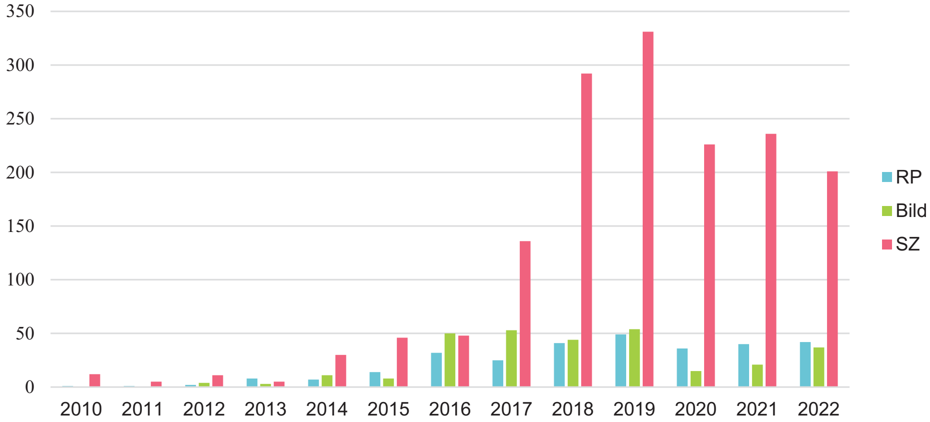


Figure 11.3: Annual number of articles on AI and algorithms published in selected newspapers.

the sentiment tone of the text. We used the “NRC” dictionary developed initially by Mohammad and Turney (2010) for the analysis since it allows us the flexibility of checking the text in German language, something that is not offered in other three dictionaries in the syuzhet package.

We wanted to examine the overall emotional tendency of the text and so we opted for calculating the mean sentiment score per article. On average, a score close to –1.0 indicates a negative sentiment of the text while a score of +1.0 indicates a positive sentiment. A score of 0 or close to zero usually means that the text is neutral. The score generated by the software was also manually checked – each member of our team randomly selected ten articles to read and compare the score to that generated by R. In almost all the cases, the manual score and that generated by R was similar.

11.3.3.1 Sentiment Score

Our results show that majority of the articles (around 86% of all articles) in all three newspapers have a positive sentiment while a very small percentage (around 1.5% of all articles) can be classified as neutral. Figure 11.4 summarises the results across different newspapers.

The highest percentage of articles with a negative tone was found in Bild (12.67%), followed by SZ (11.91%) and RP (9.69%). On the other hand, RP led the way in articles with an average positive tone (88.3%), closely followed by Bild and SZ (both at 86%). The share of neutral articles was very low. The highest number of articles were published in SZ (2.08%), followed by RP (1.99%) and Bild (1.33%). These results are relatively different from those of Fischer and Puschmann (2021) who found no articles with a negative tone for Bild and a larger percentage of neutral articles for SZ. Since

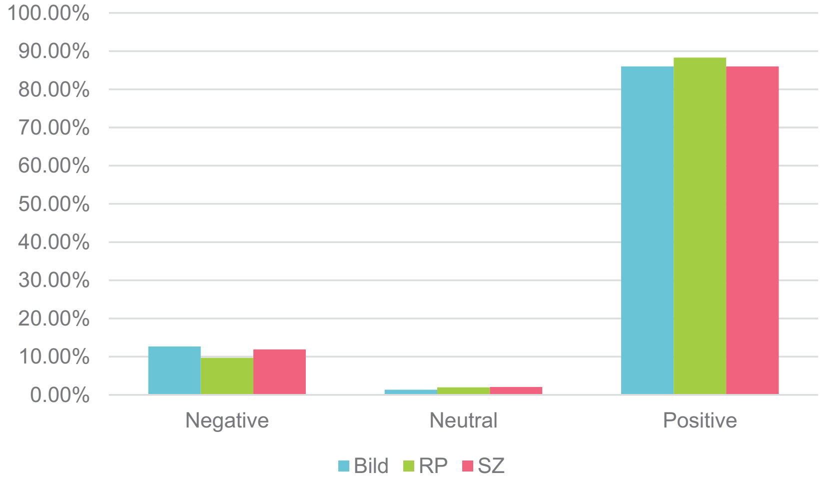


Figure 11.4: Percentage of articles with negative, neutral and positive sentiment scores per newspaper (authors' own calculation).

the report by Fischer and Puschmann (2021) does not explain what dictionary and/or method it is using to calculate the sentiment score, it is difficult to clarify the reasons behind such a big difference.

In addition to looking at the sentiment score as a percentage of articles per newspaper, we also calculated the average sentiment score for each newspaper. We found that on average, all three newspapers have a sentiment score of between 0.37 and 0.41, indicating that the German newspapers have reported somewhat positively on AI and algorithms over the years.

11.3.4 AI and Consumers

Apart from checking the sentiment score for newspaper articles on AI and algorithms, we also wanted to examine the extent to which newspapers in our sample related AI and algorithm topics to consumers, consumer protection and consumer-associated application or issues. For this purpose, we defined a new set of keywords for consumers and consumer-related issues based on our understanding of the literature (see Table 11.2) and searched for them in our sample of 2240 articles. Once again, we used the same criteria as earlier: if the keyword is appearing more than once in the text, we consider it, otherwise the text is dropped.

Table 11.3 summarises our results. In general, we observe that around 30% of the total articles on AI, for each newspaper, discussed consumers or consumer-related issues. We also checked for words like discrimination and fairness in connection with consumers in the article text and found that it is not that intensively discussed in the

Table 11.2: List of keywords for consumer-related article search (authors' own compilation).

Consumer keywords
Diskriminierung
diskriminieren
Fairness
fair
Intransparenz
Käufer
Konsumenten
Konsument
Kunden
Nachfrage
Nutzer
Transparenz
Verbraucher
Verbraucherentscheidungen
Verbraucherschutz
Vertrauen

newspapers. A comparison of the average sentiment score of the articles that discuss consumers and AI with the overall data sample shows that it does not change that much. This indicates that newspapers do not treat themes like consumers, consumer protection, fairness and discrimination any differently than other themes and do not specifically focus on it as well in their reporting.

Table 11.3: Sentiment score for AI and consumer-related articles (authors' own calculation).

Newspaper	Articles about AI and algorithms	Sentiment score	Articles about AI, algorithms and consumers	Sentiment score
SZ	1589	0.38	507	0.29
Bild	300	0.37	90	0.43
RP	351	0.41	108	0.36

11.4 Discussion of Results and Public Opinion

The main conclusion that we can draw from our results so far is that newspapers in Germany are relatively optimistic about AI and algorithm-related topics and present a fairly positive picture to their subscribers as well. However, studies have indicated that negative news have a higher impact on readers than positive news leading to feelings of distrust (McIntyre & Gibson, 2016), emotional instability and apprehension about potential harm to one self (Aust, 1985; Mclntyre & Gibson, 2016). Therefore, the important question is that how does the public in Germany perceive these themes? Does positive reporting have an impact on acceptance of these technologies among the public? We examined different public polls and found that public opinion is relatively mixed but is more inclined in a positive direction. For example, in a survey conducted by TÜV-Verband in 2019, the participants were asked *how do they feel about AI?* (TÜV-Verband, 2021). Almost 46% of the participants responded positively (TÜV-Verband, 2021). When the survey was repeated again in 2021, the positive response increased to almost 51% (TÜV-Verband, 2021). Similarly, Bertelsmann Stiftung conducted a poll in 2022, asking "*What do you think of when you hear the term 'algorithm'/'artificial intelligence'*?" (Overdiek & Petersen, 2022). More than 50% of the participants answered that they link it to *progress* and *accuracy*. However, many of them (around 51%) also thought of it as *scary* and *dangerous*. When it comes to relaying on AI to get information or to share personal information with AI, the response of the public changes as can be seen by the poll conducted by KPMG in 2020 where 39% of the respondents were unwilling to rely on AI for information purposes (Gillespie, Lockey & Curtis, 2021).

Altogether, it is obviously difficult to prove that positive sentiments in the newspapers have resulted in a more positive attitude of the public for some aspects of AI and algorithmic systems. However, it is clearly visible from our analysis that the general attitude of the newspapers as well as the public is accepting of the new technologies. It would however be more beneficial if newspapers discuss AI in relation to its impact on consumers more often. If consumers have more information about the usage and implementation of AI and algorithms, it is possible that the *scary* and *dangerous* sentiment that was observed in some polls can be reduced.

We understand that our research is not without limitations. To begin with, it is difficult to identify one (or a set of defined) terms for artificial intelligence and algorithms that is used by newspapers which makes it difficult to select relevant keywords as discussed in Section 12.2. However, we considered a diverse set of keywords to make the analysis as comprehensive as possible and would like to extend our research in the future by including context-based sentiment analysis. We are also aware of the fact that the three chosen newspapers might not be representative of overall media of Germany but we do strongly believe that the newspapers in our sample do represent a certain stratum of German population and their opinion. That being said, our research results can add to the literature discussing the role of media in general and newspapers in particular to improve the overall impression of new technologies like AI on the public.

Bibliography

Armbruster, A. (2017). Elon Musk und Co. warnen vor Killer-Robotern. *Frankfurter Allgemeine Zeitung*. https://www.faz.net/aktuell/wirtschaft/kuenstliche-intelligenz-elon-musk-warnt-vor-killer-robotern-15161436.html?profillogin (accessed 23 November 2022).

Aust, C. F. (1985). Judgments of well-being after exposure to televised bad and good news (Master's thesis). http://eric.ed.gov/?id=ED322547 (accessed 18 April 2023).

Axel Springer. (2022). Reach of the newspaper Süddeutsche Zeitung in Germany from 2003 to 2022 (in million readers) [Graph]. In: Statista. https://www.statista.com/statistics/1114895/suddeutsche-zeitung-reach-in-germany/ (accessed 5 September 2023).

Bild. (2022). Nothing moves Germany like BILD – The biggest entertainment and new portal. https://www.mediaimpact.de/en/portfolio/bild-digital-2 (accessed 29 August 2023).

Bitkom Research. (2022). Digitalisierung der Wirtschaft. https://www.bitkom.org/sites/main/files/2022-07/Bitkom-Charts%20Digitaliserung%20der%20Wirtschaft%2020%2006%202022_final_neu.pdf (accessed 23 November 2022).

Bundesverband Digitale Wirtschaft e.V. (2022) KI-Monitor 2022. https://www.bvdw.org/wp-content/uploads/2023/08/KI-Monitor_2022.pdf (accessed 6 September 2023).

Bundesverband Digitalpublisher und Zeitungsverleger. (2022). Reichweite und Auflagengrößen. https://www.bdzv.de/alle-themen/marktdaten/reichweite-und-auflagengroessen (accessed 5 September 2023).

Castelo, N., Bos, M. W., & Lehmann, D. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *144*(1), 114–126. https://doi.org/10.1177/0022243719851788.

Chuan, C. H., Tsai, W.-H. S., & Cho, S. Y. (2019). Framing Artificial Intelligence in American Newspapers. Association for the Advancement of Artificial Intelligence. https://www.aies-conference.com/2019/wp-content/papers/main/AIES-19_paper_162.pdf (accessed 23 November 2022).

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology*: *General*, *144*(1), 114–126.

Dotson T. (2015). Technological determinism and permissionless innovation as technocratic governing mentalities: Psychocultural barriers to the democratization of technology. *Engaging Science, Technology, and Society*, *1*, 98–120.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A., & Anderson, B. W. (2001). Predicting misuse and disuse of combat identification systems. *Military Psychology*, 13, 147–164. https://doi.org/10.1207/S15327876MP1303_2.

Fast, E., & Horvitz, E. (2016). Long-term trends in the public perception of artificial intelligence. AAAI. https://doi.org/10.48550/arXiv.1609.04904.

Fast, N. J., & Jago, A. S. (2020). Privacy matters . . . or does It? Algorithms, rationalization, and the erosion of concern for privacy. *Current Opinion in Psychology*, *31*(1), 44–48.

Fenneman, A., Sickmann, J., Pitz, T., & Sanfey, A. G. (2021). Two distinct and separable processes underlie individual differences in algorithm adherence: Differences in predictions and differences in trust thresholds. *PLOS ONE*, *16*(2):e0247084. https://doi.org/10.1371/journal.pone.0247084.

Fischer, S., & Puschmann, C. (2021). Wie Deutschland über Algorithmen schreibt. Bertelsmann Stiftung. https://www.bertelsmann-stiftung.de/fileadmin/files/user_upload/Diskursanalyse_2021_Algorithmen.pdf (accessed 23 November 2022).

Garvey, C. (2018). A framework for evaluating barriers to the democratization of artificial intelligence. *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence* (AAAI-18). Palo Alto, CA: AAAI Press, 8079–8080.

Garvey, C., & Maskal, C. (2019). Sentiment analysis of the news media on artificial intelligence does not support claims of negative bias against artificial intelligence. *OMICS*, *23*(0). https://doi.org/10.1089/omi.2019.0078.

Gillespie, N., Lockey, S., & Curtis, C. (2021). Trust in Artificial Intelligence: A Five Country Study. The University of Queensland and KPMG Australia. https://home.kpmg/au/en/home/insights/2021/03/artificial-intelligence-five-country-study.html (accessed 15 January 2023).

Handelsblatt. (2021). Zwei Tote bei Unfall mit wohl fahrerlosem Elektroauto. https://www.handelsblatt.com/unternehmen/industrie/tesla-zwei-tote-bei-unfall-mit-wohl-fahrerlosem-elektroauto/27108214.html (accessed 23 November 2022).

Hossain, A., Karimuzzaman, M., Hossain, M. M., & Rahman, A. (2021). Text mining and sentiment analysis of newspaper headlines. *Information, 414*. https://doi.org/10.3390/info12100414.

Illner, U. (2022). *PMG Presse Monitor*. Künstliche Intelliganz: Wohin geht der Langzeittrend?. https://www.pressemonitor.de/blog/ki-kuenstliche-intelligienz-medienanalyse/ (accessed 23 November 2022).

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1). https://doi.org/10.1177/2053951718756684.

Jockers, M. (2020). Introduction to the Syuzhet Package. https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html (accessed 8 August 2022).

McIntyre, K. E. & Gibson, R. (2016). Positive news makes readers feel good: A “silver-lining” approach to negative news can attract audiences. *Southern Communication Journal*, *81*(5), 304–315. https://doi.org/10.1080/1041794X.2016.1171892.

Mohammad, S. M. & Turney, P. D. (2010). Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pp. 26–34, Los Angeles, California. http://saifmohammad.com/WebPages/lexicons.html (accessed 25 November 2022).

Nader, K., Toprac, P., Scott, S., & Baker, S. (2022). Public understanding of artificial intelligence through entertainment. *AI & Society*. https://doi.org/10.1007/s00146-022-01427-w.

Nelkin, D. (1987). *Selling Science: How the Press Covers Science and Technology*. New York: W.H. Freeman.

Nelkin, D. (1998). Scientific journals and public disputes. *Lancet*, *352*, S25–S28. https://doi.org/10.1016/S0140-6736(98)90297-5.

Ouchchy, L., Coin, A., & Dubljevic, V. (2020). AI in the headlines: The portrayal of the ethical issues of artificial intelligence in the media. *AI & Society*, *35*, 927–936. https://doi.org/10.1007/s00146-020-00965-5.

Overdiek, M., & Petersen, T. (2022). Was Deutschland über Algorithmen und Künstliche Intelligenz weiß und denkt. Bertelsmann Stiftung. https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/was-deutschland-ueber-algorithmen-und-kuenstliche-intelligenz-weiss-und-denkt-all (accessed 15 January 2023).

Ozgun, B., & Broekel, T. (2021). The geography of innovation and technology news – An empirical study of German news media. *Technological Forecasting and Social Change*, 167. https://doi.org/10.1016/j.techfore.2021.120692.

Sackmann, C. (2018, October 12). Künstliche Intelligenz erachtet Bewerbungen von Frauen als minderwertig – Amazon muss reagieren. *FOCUS online*. https://www.focus.de/finanzen/boerse/die-tuecken-der-intelligenz-amazon-schaltet-ki-ab-die-bewerbungen-von-frauen-als-minderwertig-erachtete_id_9741890.html (accessed 23 November 2022).

Shead, S. (2018, December 8). Tech journalists locked out at top AI conference. *Forbes*. https://www.forbes.com/sites/samshead/2018/12/08/tech-journalists-locked-out-at-top-ai-conference/ (accessed 23 November 2022).

Statista. (2022). https://www.statista.com/statistics/1114895/suddeutsche-zeitung-reach-in-germany/ (accessed 29 August 2023).

TÜV-Verband. (2021). *Sicherheit und Künstliche Intelligenz*. https://www.tuev-verband.de/?tx_epxelo_file[id]=856779&cHash=1af8a3f0e6c845423fdd637c8dbcd080 (accessed 15 January 2023).

Vicsek, L. (2011). Costs and benefits of stem cell research and treatment: Media presentation. *Science Communication*, *33*(3), 309–340. https://doi.org/10.1177/1075547010389820.

Zeit Online. (2018, March 18). Ermittlungen gegen Wahlkampfhelfer von Donald Trump. *Zeit*. https://www.zeit.de/digital/datenschutz/2018-03/cambridge-analytica-ermittlungen-datenschutzverletzungen-us-wahlkampf (accessed 23 November 2022).

Zhang, D., Maslej, N., Brynjolfsson, E., Etchemendy, J., Lyons, T., Manyika, J., Perrault, R. (2022). *The AI Index 2022 Annual Report*. Stanford University. https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf (accessed 7 September 2023).

Simone Roth, Medina Klicic

Chapter 12
Generational Differences in Framing for Social Robot Usage Intention from a Consumer Behaviour Point of View

Abstract: An increasing number of social robots exist globally, but are not yet used in everyday life. Literature-based and with an exploratory study, this chapter outlines options to build on existing associative structures in the minds of consumers and uses the framing approach to improve attitudes and increase the intention to use social robots. Firstly, a definitional classification of what exactly makes a robot social is presented. Secondly, generation-specific differences in attitudes and intentions to use are determined. Thirdly, the framing approach for increasing attitude, including behavioural intention is discussed. Subsequently, a series of interviews were used to investigate the associative structures with regard to social robots, generational differences in association networks and the use of the framing approach. The application field in the study is education. Findings indicate generational differences and risky choice, attribute and goal frames have been developed, upon which further research can be based.

Keywords: social robots, generations, framing

12.1 Introduction

A future in which autonomous robots assist us in every aspect of our daily lives is not only possible but very likely. In recent years, social robots have begun to be seen as one of the top emerging technologies. In 2019, the World Economic Forum ranked social robots as number 2 on their top 10 list, marking how droid friends and assistants are entering more deeply into our societies and lives (Hadzilacos et al., 2019). However, “there are still miles to go before robots become a regular feature within our social spaces” (Henschel, Laban & Cross, 2021, p. 9). Estimates from the International Federation of Robots show continuous growth rates for sales of service robots (Müller, Graaf & Pfeiffer, 2022). Although robots are found, for example in the hospitality sector, to make, sell and serve coffee (e.g., Nestle) and free up staff time (Kotler, Kartajaya, &

Simone Roth, Institute of Economics, Ruhr West University of Applied Sciences, Mülheim, e-mail: simone.roth@hs-ruhrwest.de
Medina Klicic, Institute of Economics, Ruhr West University of Applied Sciences, Mülheim, e-mail: medina.klicic@hs-ruhrwest.de

https://doi.org/10.1515/9783110790320-012

Setiawan, 2021), the absolute number sold worldwide in the main application areas is rather low. Therefore, the aim of this chapter is to bridge the gap between already existing off-the-shelf social robots and the lack of use of social robots in daily life.

The chapter is structured as follows: After setting a frame of reference that classifies a robot as social, a generational approach is used to understand existing associations and expectations consumers have built up about social robots. Additionally, differences in consumers' attitudes towards social robots from a sociocultural perspective are reviewed. Based on this, to reach the aim of this chapter, a new perspective to increase the intention to use social robots will be drawn from the behaviour science approach of framing by developing context-specific frames (Kahneman, 2012; Tversky & Kahneman, 1981). Further, it is proven that the application field and context are important for the research project (e.g., Han & Conti, 2020) and therefore the application in focus will be the educational context and specifically education in universities.

12.2 Classification of Social Robots

There is a multitude of different approaches to define and classify social robots (Bendel, 2021a), and it can be stated that the definitions of social robots are heterogeneous in the scientific literature, much like the understanding of what specifically makes them social (Henschel, Laban & Cross, 2021).

The definition by Duffy (2003) reflects the multifaceted nature of social robots: "There is a two-way interaction between a robot and that which it becomes socially engaged with, whether another robot or a person. It is in embracing the physical and social embodiment issues that a system can be termed a 'social robot'" (Duffy, 2003, p. 178). In an overall framework for social robotics by Baraka, Alves-Oliveira and Ribeiro (2020), other relevant dimensions are stated next to physical and social embodiment. Accordingly, relevant dimensions for social robots are (a) related to the robot itself, namely appearance, social capabilities and autonomy/intelligence, (b) related to the interaction, namely proximity and temporal profile and c) related to the context, namely robot relational role and purpose/application area (Baraka, Alves-Oliveira & Ribeiro, 2020). To set a common ground, the framework of Baraka, Alves-Oliveira and Ribeiro (2020) is used with a focus on dimension (a).

In a very broad form, social robots can be described as "sensorimotor machines created to interact with humans and animals" (Bendel, 2021b, p. 1). This is why, physical appearance is a first dimension, among other relevant dimensions, which can be varied in terms of definition (for an overview, see Sarrica, Brondi & Fortunati, 2020). This first decisive factor, physical appearance, influences people's perception, attitude and behavioural intention, as demonstrated in several studies (see meta-analysis by Blut et al., 2021). Regarding the appearance and physical embodiment, a distinction can be made between bio-inspired (e.g., human- or animal-inspired), artefact-shaped (e.g., robots resembling man-made objects or those that are imaginary) and functional robots (e.g., drones) (Baraka, Alves-Oliveira & Ribeiro, 2020). The second decisive factor is how social robots relate to humans. So, the question is, what is behind the term "social"? Despite their heterogeneity, the definitions extracted from the scientific literature related to social capabilities of robots show similarities. Their mutuality is found in the concept of interaction. The interaction shapes the relationship between machine and human being. Therefore, perceiving and responding to environmental cues, engaging in social interactions, communicating, cooperating, learning, making decisions and performing actions are mentioned qualities of social interaction (Sarrica, Brondi & Fortunati, 2020). This is determined by action and reaction (Bendel, 2021b). In particular, communication as a crucial part of interaction can be done one-way, meaning the user sends a message (commands, questions) to the social robot. The robot's answer follows in the mere act of executing according to the instructions, for example a vacuum cleaner robot. Alternatively, a two-way communication in which the social robot is autonomously reacting and responding to its counterpart is possible, for example a robot serving in a restaurant. When communicating with a living being, the social robot is equipped with a function that enables natural language as well as gestural and mimical ability, allowing normal dialogues to take place. To allow social interaction and communication, customer-facing interfaces are requirements for the social robot (Kotler, Kartajaya & Setiawan, 2021).

Applying the proposed definitional dimensions of social robots to existing (social) robots in the market,[1] a classification can result in the corresponding matrix (Figure 12.1). The matrix considers physical embodiment (artefact-shaped and bio-inspired robots) and communication (one-way- and two-way-communication) as a part of social interaction.

To summarise the classification of social robots: Only physical robots with social capabilities and a certain autonomy are considered in this chapter. In the matrix, these robots can be found under two-way communication (y-axis), including artefact-shaped and bio-inspired robots (x-axis). Excluded from this chapter are software robots like chatbots, voice bots, voice assistants or virtual assistants, social bots, as well as robot process automation (RPA).[2] Additionally, industrial robots and robots only able to receive commands/messages from humans are not included.

1 An overview of more than 200 robots used in many markets and categories as well as prototypes for research can be found in collections such as the ABOT database (Phillips et al., 2018).

2 For a definition for software robots see Bendel (2021a) as well as for RPA Kotler, Kartajaya & Setiawan (2021). In RPA, the virtual robot performs computer work as a human would, following specific guidelines.

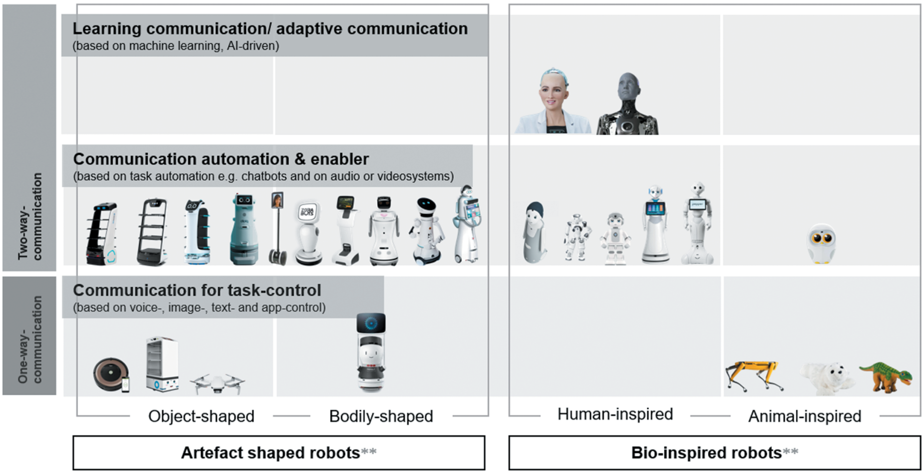


Figure 12.1: Classification matrix for social robots in the market. (compiled by authors.)

12.3 Generational Differences in Attitudes Towards Social Robots

In addition to the dimensions of social robots, other factors – cultural, societal, and individual – also influence the complex construct of attitudes. Based on the so-called affect-behaviour-cognition model (ABC model), attitudes consist of the three components mentioned above: *affect* refers to the feelings of a consumer related to an attitude object. *Behaviour* describes actions the consumer takes towards the object including the behavioural intentions. Lastly, *cognition* summarises what consumers believe to be true about the attitude object (Solomon, 2017). Additionally, interrelationships exist between attitude components. In conclusion, the attitude towards social robots can only be determined by understanding these components and their interrelationships. After a systematic review of 97 studies, Naneva et al. (2020) found consistency in affective, cognitive and general attitudes towards social robots; the findings suggest that consumers have slightly positive attitudes towards social robots and are willing to interact with them. Noteworthy, the literature review reveals an ambiguity of attitudes towards social robots, meaning negative and positive attitudes towards social robots exist at the same time (e.g., Stapels & Eyssel, 2021).

In the following, we focus on the influence of individual characteristics on attitude (components). Studies on gender, fear and previous experience on attitude support those influences (Duffy, 2003). Furthermore, there are also studies that examine the intention to use social robots (behavioural component of attitudes), for example studies regarding the impact of age on willingness to use robots (Broadbent, Stafford & Mac-

Donald, 2009; Kuo, Wu & Deng, 2009) or gender influences on scepticism about interacting with robots and likeliness to use them (Abel et al., 2020; Athanasiou et al., 2017; De Keyser & Kunz, 2022; Graaf & Ben Allouch, 2013;). Lastly, a recent meta-analysis by Blut et al. (2021), summarised individual factors and their influence on intention to use and interrelations with anthropomorphism. Central moderators (service type, perceived skill of the robot, etc.) and further mediators (robot-related such as intelligence and safety) were also examined (Blut et al., 2021). To structure the complexity of the individual factors influencing the attitudes towards social robots, the generational perspective can be a simplifying entrance point. Every generation is shaped by a different sociocultural environment and life experiences and those are particularly formative in the years of a person's adolescence (Duffy, 2021; Kotler, Kartajaya & Setiawan, 2021). Significant experiences can be events like the fall of the Berlin Wall, environmental catastrophes, but – important in this context – also technological tools (Scholz, 2014). This means that generations can have different attitudes because of different socialisation conditions. The so-called cohort effect states that attitudes will remain distinct from other cohorts even as they age (see Table 12.1) (Duffy, 2021).

Table 12.1: The generational split.

Generations	Traditionalists/ pre-war[1]	Baby Boomer[1]	Gen X[1]	Gen Y[1]	Gen Z[1]	Gen Alpha[2]
Born	Before 1945	1945–1965	1966–1979	1980–1995	1996–2010	As of 2011

Note: 1. (Duffy, 2021), 2. In accordance to (Kotler, Kartajaya & Setiawan, 2021).

Technological accelerations and changes contribute to a growing disconnection between cohorts (Duffy, 2021). Generation X experienced the launch of the first personal computers with cash register drives on the market, while Generation Y was growing up when Google was introduced and lives turned digital (Roth & Nazemian, 2019). A source of difference between generations was identified by Duffy (2021) in terms of interaction and communication through technology. The adoption of the smartphone as one central technology of modern life is an example of the differences in smartphone use by the different cohorts (Duffy, 2021). Generation Z has known little else than life with a smartphone, while Generation Y's experiences varied over the course of their adult lives. In contrast, only seven in ten British Baby Boomers own a smartphone (Duffy, 2021), meanwhile Generation Alpha, due to "screen fatigue" from the last COVID years, is moving away from technology (Buckle, Moran & Trivonova, 2022). The intensity of smartphone use, as well as the types of things done online, varies significantly between generations. In particular, social media use and chosen social media platforms are differentiating aspects visible in longitudinal studies (Duffy, 2021). To conclude this part, it is important to consider, that the generational perspectives on technology and communication might differ but this is certainly not true for every individual in a generation and not to the same extent (Deal & Levenson, 2016; Havens, 2015). Moreover, there is no clear demarcation between generations in the scientific literature (Deal & Levenson, 2016; Havens, 2015; Kotler, Kartajaya & Setiawan, 2021). Nevertheless, despite all deviations, a constant core is visible and this perspective is considered to be useful for influencing attitudes towards social robots.

12.4 Framing Social Robot Use Intention

In the following part, a new perspective to influence attitudes towards a social robot will be explored using the approach of framing.

The basic idea is that the willingness to use a social robot can be influenced by the wording in which the social robot is introduced and presented. A framing effect is the unjustified influence of a change in the wording of a decision problem on beliefs and preferences (Kahneman, 2012). Kahneman (2012) explained what can happen in one's associative machinery while understanding sentences, framing an event in different ways: the two sentences "Italy won" versus "France lost" are both descriptions of the result of Italy playing against France in the World Cup final 2006 (Kahneman, 2012). However, the associations that are evoked in the mind of the reader are different. In statement one, the evoked associations are connected to what the Italian team did to win, and the second sentence evokes associations with what caused the loss of the French team. The differently aroused associations might also evoke different reactions (Kahneman, 2012). The framing approach was used in a number of fields in economics already. One of the first examples dates back to the introduction of credit cards in the 1970s. As the use of credit cards as a form of payment became more and more accepted, retailers wanted to compensate for the fees they had to transfer to the credit card companies. So, if a customer wanted to use a credit card instead of paying cash, they wanted to add a "credit surcharge." The credit card companies tried to prevent this by influencing regulations but created a fallback option with a focus on language (i.e., a different frame): if a retailer wants to charge different prices depending on the kind of payment, the credit card price should be framed as "normal" and a "cash discount" given for payments in cash (rather than "framing" it as a credit surcharge). (Thaler, 2021). It is precisely this approach that is of interest below in connection with the introduction of social robots. A recent publication explained that framing effects are intensified by the fact that humans tend occasionally "to be somewhat mindless, passive decision makers. Few of us bother to check to make sure reframing the decision we face would produce a different answer" (Thaler, 2021, p. 40).

Different types of frames can be distinguished, depending on whether positive or negative terms are used to effect judgements and decisions. A typology and critical appraisal can be found, for example in Levin, Schneider and Gaeth (1998). The typology results from the fact that the frames influence different underlying processes through different operationalisations. Accordingly, the frames have a different valence (Levin, Schneider & Gaeth, 1998). Basically, three types of frames can be distinguished (Levin, Schneider & Gaeth, 1998): the first type, the *standard risky choice framing,* goes back to the original work of Tversky and Kahneman. They created frames showing outcomes of a potential choice and the described choices differ in the level of risks (Tversky & Kahneman, 1981). The constituent feature of risky choice framing is the possibility of choosing between a safe and an unsafe (risky) option in a decision problem. It was demonstrated that discrete choices between a risky and a riskless option of equal expected value depended on whether the options were described in positive terms (i.e., lives saved) or in negative terms (i.e., lives lost). So, the outcomes could be framed as "Win" or "Loss" (Tversky & Kahneman, 1981).

The second type is the so-called attribute framing, which affects the evaluation of an object or event characteristics (Levin, Schneider & Gaeth, 1998). In attribute framing, a single attribute within any given context is the subject of the framing manipulation. For example, the choice can be a yes/no judgment (Would you be in favour of the programme?) or participants are asked to give ratings of favourability (Rate the programme on a scale from bad to good) (Levin, Schneider & Gaeth, 1998). In this case, providing an evaluation of the favourability of either acceptance or rejection would specify the choice (Levin, Schneider & Gaeth, 1998). Studies showed that the same alternative was rated more favourably when described positively than when described negatively. A number of studies have used this type of manipulation to examine the evaluation of subjects, such as job placement programs in industry project teams (an overview can be found in Levin, Schneider & Gaeth, 1998).

The third type is the goal framing, which affects the persuasiveness of communication. It is intended to improve the evaluation of a situation or behaviour (Levin, Schneider & Gaeth, 1998). In particular, the subject may be framed to focus attention on its potential to provide a benefit or gain (positive frame) or on its potential to prevent or avoid a loss (negative frame) (Levin, Schneider & Gaeth, 1998). A well-known example of a goal-framing effect was documented by Meyerowitz and Chaiken (1987). They showed that women were more willing to engage in breast self-examination when presented with information emphasising the negative consequences of not doing so than when presented with information on the positive consequences of such an exam. It is suggested that the effect occurs because of a negativity bias in processing information. This implies that negative information has a systematically stronger impact on judgment than objectively equivalent positive information (Meyerowitz & Chaiken, 1987). This framing effect has also been found in a number of studies (Levin, Schneider & Gaeth, 1998).

All three types of framing can be a useful approach for influencing the intention to use social robots. To create according to frames, it is needed first to explore and identify associations with regard to social robots and based on this, develop context-specific frames to increase the willingness to use social robots.

12.5 Empirical Findings

An exploratory study was conducted to gain insights into associations about social robots in an educational context[3] and intergenerational differences, and to indicate a starting point for developing potential frames for future research.

12.5.1 Methodology

The aim was to investigate the following questions: *(1) Which associations exist with regard to social robots and what differs between generations? (2) Which tasks, roles and skills are expected in education by generations? (3) Which factors need to be considered in relevant frames?* The methodology used was exploratory research in the form of personal semi-structured interviews. The study included all age groups but focused on generation X, Y and Z. A convenience sample approach was used. A total of n = 17 interviews were conducted in January 2023 (see Table 12.2). The sample showed different genders, educational backgrounds, ages and tech savviness. The measuring instrument was a semi-standardised questionnaire consisting of four sections: (a) introduction, (b) main section on social robots, (c) technology and communication and (d) socio-demographic data. The interviews were content analysed to determine common structures for specific research objectives.

12.5.2 Results of the Preliminary Study

Starting the analysis, the socio-demographic data were summarised (see Table 12.2), respecting the generational cohort split. The participants equally represent Generation X, Y and Z (by age), while Baby Boomers and Generation Alpha were represented by n = 1. As outlined in Section 13.3, the generations differ in tech savviness and use of social media, which is also included.

Looking at the structure of the study participants, the degree of technical savvy seems to correspond to the generational understanding, even if the data only reflects a tendency. Regarding social media, younger generations seem more likely to state that social networks are part of their daily activities than older generations. In terms of activity level, younger generations seem to be heavier users but tend to be more passive (observing and consuming). The results therefore confirm a similar pattern as in a large number of other studies over generations (Duffy, 2021). For the upcoming

3 The educational context requires an institution whose primary focus is to provide education. In this research project, the educational study was applied to a university of applied sciences.

Table 12.2: Social-demographics, social media usage and tech savviness. (compiled by authors.)

Generation	Baby Boomer	Gen X	Gen Y	Gen Z	Gen Alpha
Sample n=17	n=1	n=5	n=5	n=5	n=1
Age (respondent)	r17=57	r12=43; r13=49; r14=51; nr5=53; r16=56	r7=27; r8=27; r9=32; r10=34; r11=36	r2=14; r3=16, r4=20; r5=21; r6=22	r1=10
Pew digital savviness classifier (PDS)[4]	Somewhat digitally savvy (n=1)	Very digitally savvy (n=1); somewhat digitally savvy (n=4)	Very digitally savvy (n=4); somewhat digitally savvy (n=1)	Very digitally savvy (n=3); somewhat digitally savvy (n=2)	Somewhat digitally savvy (n=1)
Usage pattern of social media (top two boxes): daily use and active participation	n=0 n=0	Daily use (n=3); active participation (n=2)	Daily use (n=4); active participation (n=0)	Daily use (n=5); active participation (n=0)	n=0 n=0

analysis on association networks, Baby Boomer as well as Generation Alpha are excluded due to the small number in the sample.

An exploration into the associative networks towards social robots from a generational perspective revealed the following answers to the main research questions:

(Q1) Which associations exist with regard to social robots and what differs between generations? According to activation models of memory,[5] incoming information is stored in an associative network that contains many bits of information and combines nodes and links (Solomon, 2017). Each node represents a concept related to social robots and is connected by links. These nodes can be an attribute, a specific brand or a related product. Associations like this can be visualised as a complex spider web with pieces of data. In this research context, n = 15 respondents are reflected in the associative network (see Figure 12.2) and are drawn from answers to an unprompted, open-ended question about the associations (part B, questionnaire).

4 The classifier is based on responses to two questions measuring the level of use and comfort with digital technology.

5 According to (non-modal) theories of knowledge, our knowledge is organised in schemata, which can be represented as semantic networks. A schema is an organisational form of knowledge that like any meaningful knowledge, can be represented as a semantic network. If one or some element(s) of an internalised schema is/are addressed, there is a high probability that the other stored elements belonging to this schema will also be recalled (even if these properties are not stored in the scheme. For further details in memory and activation theory in schemes see Kroeber-Riel, W. & Gröppel-Klein, A. (2019).)

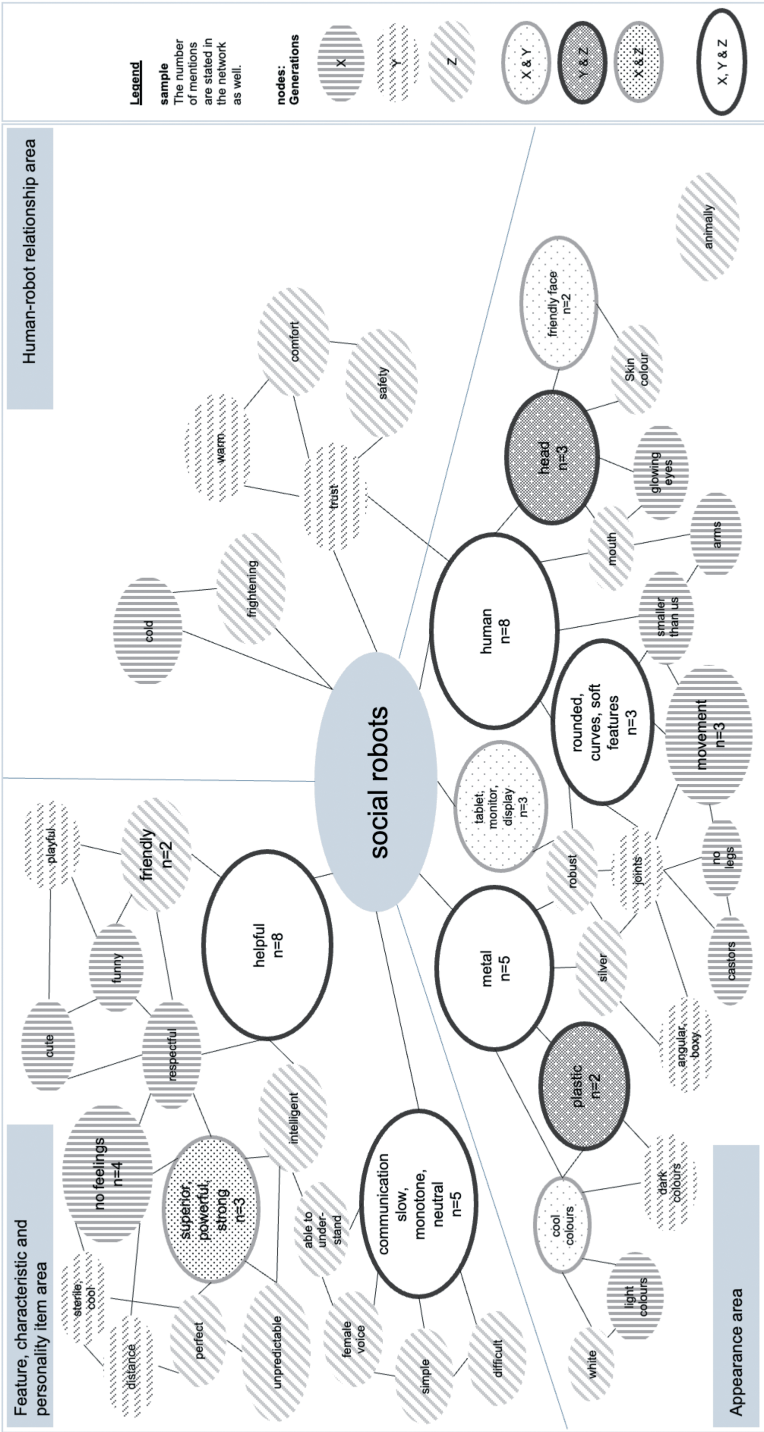


Figure 12.2: Associative network of social robots clustered in content areas. (compiled by authors.)

The associative network is visualised and structured in three content areas with regard to social robots: (1) appearance, (2) features and characteristics of the robot and (3) human–robot relationship. Associations with the (1) appearance of the social robot were multifaceted. This seemed to be consistent across generations. All generations named associations with human (n = 8) appearance, with additional single specifications, such as “head,” “mouth” and “arms,” as well as the metallic physique (n=5), permeating the associations of many respondents across generations. Furthermore, participants (n=3) described the robot’s appearance as “round” and “soft.” In area (2), features and characteristics, the characteristic “helpful” (n=8) emerged as a clear commonality among the generations. The human–robot relationship (3) was the least detailed. Generation Z had the highest number of associations in this field (n=3), besides all the positive aspects, such as “comfort” and “safety,” negative associations, such as “frightening,” were also expressed. It is remarkable that these associations mirror the ambiguity in attitudes towards social robots described in Section 13.3. Excluded from the graphic are tasks and application fields.

In the following, special consideration was given to the dimension’s appearance and communication in the educational context. To get a deeper understanding of associations connected to appearance, the respondents were given a list of 17 social robots and asked to indicate which they preferred. Based on the findings, the robot AlicePro was chosen most often. However, a closer look revealed differences across generations (see Figure 12.3).

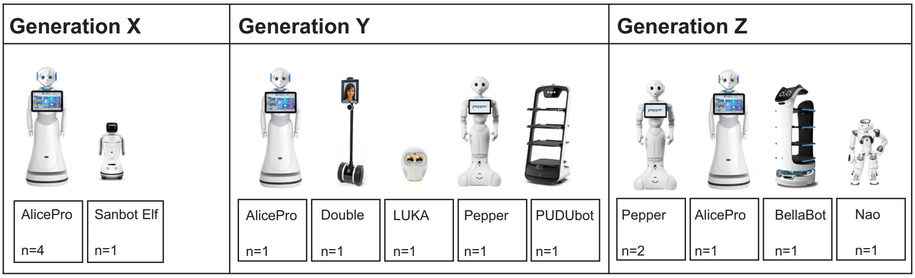


Figure 12.3: Social robot preference by generations (number of participants). (compiled by authors.)

As a reason for the choice, Generation X stated that the robot was “sympathetic” and had a “positive appearance.” According to the respondents, this was due to the “face [that] looks human” and the external appearance characteristics, such as “female” and “good size.” Generation Y was quite diverse in justifying preferences ranging from “sympathetic appearance” and “human-like (face, size)” to “woman” in a way similar to Generation X. Additionally, they mentioned “taking people who are not there with you” to justify the preferred robot Double. Generation Z named aspects like “can move,” “looks the friendliest,” “humanoid features,” but, compared to other generations, also addressed aspects regarding expected functionality. For example, “[it has] in front a tablet and can show me what I want to see," "[can] talk, reach because it has two arms," "looks like it could help" or "answer questions."

(Q2) Which tasks, roles and skills in education are expected by generations? The application context plays a vital role in attitude formation (Naneva et al., 2020). Therefore, the study focused on a specific application to develop frames. Based on answers to an open-ended question about expected tasks, it became clear that Generation X highly appreciates assistance with tasks that provide physical relief. This is exemplified by statements such as "a toy" or "kitchen help." In an educational setting, respondents from this generation expected a social robot to "welcome guests," be an "information booth," "bring coffee," "serve drinks and snacks," "check tickets," "[present a] site plan of the venue" or "show presentations." Generation Y goes one step further in citing capabilities such as "enable communication," act as "voice assistant" and "remember things," as well as "call via video chat with people who cannot be there." Here, the focus is more on communication with and through the social robot. Nevertheless, this type of communication is limited to very simple tasks, which is congruent with the negatively connotated statements ("monotone," "robot in interaction cannot react immediately") in association networks. In the educational field, respondents explained that the robot can assist, in "set up," "guest support," "reception," "handing out on name tags," "handing out drinks," "giving directions," "taking exactly those people who cannot be on site," which is congruent with the expected functions. Generation Z goes further into the communicative aspect. Compared to Generation Y, more autonomy seems to be expected in the activity of the social robot. Statements such as "drive away," "reach with arms," "move head," "talk to me," "show things on the tablet that I want to see," "asking how something works" and "explanation" reinforce this conclusion. They also explained that organisational tasks could be a role of the social robot. However, aspects such as "show examples using the tablet," "doing things with arms and hands," "teaching theoretical concepts using the tablet," "could take on the teaching role," "illustrate lessons using the tablet" and "at rallies and stations" go beyond previously mentioned activities. That said, ambiguity in attitude is evident in the associations regarding communication above with "simple," "friendly voice," "female" and mirrors in parts the negative aspects like "difficult to communicate" and neutral elements like "one is understood" and the "neutral voice."

Generation Z seems to expect an active teaching role and not simply organisational support or retrieval and delivery tasks, which was the focus for Generation X. Generation Y complements the starting points of Generation X with a few aspects of assistance, such as voice assistance and telepresence. For Generation Y, a social robot is a means to an end to connect people. Finally, it needs to be mentioned that one respondent of Generation Alpha mentioned telepresence as well, but to the highest degree of autonomy, for self-presentation, to be able to actively participate in events. It needs to be stated that the qualitative study included only a certain number of respondents and therefore has limitations. Additionally, not all influencing factors could be deeply investigated in the qualitative study due to time limits. Lastly, the generational perspective needs to be further investigated.

12.5.3 Future Research

The findings provide initial starting points for the frames to be developed for social robots in education. Therefore, the focus is now on the last question: (*3) Which factors need to be considered in relevant frames?* The results revealed that the expectation and existing association network with regard to social robots differ. The following resulting ideas will be used: the frames can (a) build on existing association networks; this means that they can explicitly cover all three content areas. Additionally, the frames can (b) include role-specific descriptions of the social robot, covering a wide angle on communication in task performance. Lastly, the frames (c) can include a visual representation of a social robot. As a result of the findings, generation-specific frames to increase the intention to use social robots were developed (see Table 12.3). For each type of frame in the application to social robots, a constant core will cover all three identified association areas including role-specific decisions. Additionally, for each of the selected frames, the generation-specific findings of the existing associations with social robots can be used for adaptation. For example, in the attribute frame, a task description in a second sentence can be chosen to match the expected task of a social robot for each generation (see question "Which tasks, roles and skills in education are expected?"). In the case of Generation Z, the frame could then be

Table 12.3: Frame propositions for future research. (compiled by authors.)

Risky/choice frames	Frame 1: Would you use a social robot that offers a 10% chance to win 20 hours in preparing a written elaboration and a 90% chance to lose 4 hours in getting to know the robots' functionalities? Frame 2: Would you invest 4 hours to get to know a social robot's functionality that offers a 10% chance to win 24 hours and 90% chance to win nothing?
Attribute frame (Win/Loss, e.g., Generation X)	Imagine you are responsible for the qualification programme in higher education didactic. *Win frame*: Would you invest 5 hours of your time to learn how to use a social robot for teaching purposes, if you knew that already 20 out of 50 uses of a social robot in higher education were successfully completed? The social robot had the task to welcome students, had a site plan of the campus and was an overall information booth. *Loss frame*: Would you invest 5 hours of your time to learn how to use a social robot for teaching purposes, if you knew that already 30 out of 50 uses of a social robot in higher education were completed unsuccessfully? The social robot had the task to welcome students, had a site plan of the campus and was an overall information booth.

Table 12.3 (continued)

Goal frames	Frame: Technology advancements are supporting us in everyday life. Imagine you could use the social robot on the left (Robot AlicePro vs. Pepper) like your smartphone for the creation of a written elaboration. Generation-specific addition, e.g., Generation X: AlicePro would be able to search for literature, create suggestions for headlines or connect you to experts in the relevant field of interest. Would you be willing to use AlicePro?

further adapted in the sense of an independent, more active teacher role. This also applies to the adjustments to the goal frame.

In the next step, these frames will be tested in different generations in a qualitative and quantitative study to evaluate if the intention to use social robots can be increased in an educational context. Frames will be adapted to generations accordingly.

12.6 Conclusion

The behavioural science approach of framing constitutes an entry point for the shaping of attitudes and increasing the intention to use social robots. Every generation has different preferences and attitudes towards technical products and services. Social robots offer various dimensions to be considered (appearance, communication to socialise, etc.). The explorative study was undertaken to shed light on association networks and attitudes towards social robots. Besides existing associations, the task horizon of social robots was developed and further investigated. The resulting frames can be a starting point to influence intentions to use. In future research, other types of frames can be included as well.

Bibliography

Abel, M., Kuz, S., Patel, H. J., Petruck, H., Schlick, C. M., Pellicano, A., & Binkofski, F. C. (2020). Gender effects in observation of robotic and humanoid actions. *Frontiers in Psychology*, *11*, 797. https://doi.org/10.3389/fpsyg.2020.00797 (accessed 3 January 2023).

Athanasiou, A., Xygonakis, I., Pandria, N., Kartsidis, P., Arfaras, G., Kavazidi, K. R., Foroglou, N., Astaras, A., & Bamidis, P. D. (2017). Towards rehabilitation robotics: Off-the-shelf BCI control of anthropomorphic robotic arms. *BioMed Research International*, 2017, 5708937. https://doi.org/10.1155/2017/5708937 (accessed 26 December 2023).

Baraka, K., Alves-Oliveira, P., & Ribeiro, T. (2020). An extended framework for characterizing social robots. In: C. Jost, B. Le Pévédic, T. Belpaeme, C. Bethel, D. Chrysostomou, N. Crook, M. Grandgeorge, & N. Mirnig (Eds.), Springer Series on Bio- and Neurosystems. *Human-Robot Interaction* (Vol. 12, pp. 21–64). Springer International Publishing. https://doi.org/10.1007/978-3-030-42307-0_2 (accessed 4 January 2023).

Bendel, O. (Ed.). (2021a). Business book summary. Soziale Roboter: Technikwissenschaftliche, wirtschaftswissenschaftliche, philosophische, psychologische und soziologische Grundlagen. Springer Gabler.

Bendel, O. (2021b). Die fünf Dimensionen sozialer Roboter. In O. Bendel (Ed.), Business book summary. Soziale Roboter: Technikwissenschaftliche, wirtschaftswissenschaftliche, philosophische, psychologische und soziologische Grundlage (pp. 1–20). Springer Gabler.

Blut, M., Wang, C., Wünderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, *49*(4), 632–658. https://doi.org/10.1007/s11747-020-00762-y (accessed 27 December 2023).

Broadbent, E., Stafford, R., & MacDonald, B. (2009). Acceptance of healthcare robots for the older population: Review and future directions. *International Journal of Social Robotics*, *1*(4), 319–330. https://doi.org/10.1007/s12369-009-0030-6 (accessed 6 January 2023).

Buckle, C., Moran, S., & Trivonova, V. (2022). Gen Alpha: The real picture. https://www.gwi.com/reports/gen-alpha (accessed 2 January 2023).

De Keyser, A. de, & Kunz, W. H. (2022). Living and working with service robots: A TCCM analysis and considerations for future research. *Journal of Service Management*, *33*(2), 165–196. https://doi.org/10.1108/JOSM-12-2021-0488 (accessed 29 December 2023).

Deal, J. J., & Levenson, A. R. (2016). *What Millennials Want from Work: How to Maximize Engagement in Today's Work*. Business Books. McGraw-Hill Education – Europe.

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3–4), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3 (accessed 27 December 2023).

Duffy, B. (2021). *Generations: Does When You're Born Shape Who You Are?* Atlantic Books.

Esch, F.-R., Tomczak, T., Kernstock, J., Langner, T., & Redler, J. (Eds.). (2019). *Corporate Brand Management*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-24900-7 (accessed 30 December 2023).

Graaf, M. M. de, & Ben Allouch, S. (2013). Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems*, *61*(12), 1476–1486. https://doi.org/10.1016/j.robot.2013.07.007 (accessed 3 January 2023).

Hadzilacos, R., O'Halloran, D., Rusting, R., & Woeffray, O. (2019). *Insight Report: Top 10 Emerging Technologies 2019*. Geneva.

Han, J., & Conti, D. (2020). The use of UTAUT and post acceptance models to investigate the attitude towards a telepresence robot in an educational setting. *Robotics*, *9*(34), 1–19. https://doi.org/10.3390/robotics9020034 (accessed 27 December 2023).

Havens, J. (2015). *Us vs. Them: Redefining the Multi-Generational Workplace to Inspire Your Employees to Love Your Company, Drive Innovation, and Embrace Change* (1st edition). PH Professional Business. https://learning.oreilly.com/library/view/-/9780134195193/?ar (accessed 4 January 2023).

Henschel, A., Laban, G., & Cross, E. S. (2021). What makes a robot social? A review of social robots from science fiction to a home or hospital near you. *Current Robotics Reports*, *2*(1), 9–19. https://doi.org/10.1007/s43154-020-00035-0 (accessed 26 December 2023).

Jost, C., Le Pévédic, B., Belpaeme, T., Bethel, C., Chrysostomou, D., Crook, N., Grandgeorge, M., & Mirnig, N. (Eds.). (2020). *Human-Robot Interaction*. Springer Series on Bio- and Neurosystems. Springer International Publishing. https://doi.org/10.1007/978-3-030-42307-0 (accessed 2 January 2023).

Kahneman, D. (2012). *Thinking, Fast and Slow* (Di 1 ban). Penguin.

Kanda, T., Ŝabanović, S., Hoffman, G., & Tapus, A. (Eds.) (2018). *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM.

Kotler, P., Kartajaya, H., & Setiawan, I. (2021). *Marketing 5.0: Technology for Humanity* (1st ed.). Wiley. https://learning.oreilly.com/library/view/-/9781119668510/?ar (accessed 30 December 2023).

Kroeber-Riel, W. & Gröppel-Klein, A. (2019). Konsumentenverhalten (11th ed.). Franz Vahlen. https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=6991291 (accessed 30 December 2023).

Kuo, Y.-F., Wu, C.-M., & Deng, W.-J. (2009). The relationships among service quality, perceived value, customer satisfaction, and post-purchase intention in mobile value-added services. *Computers in Human Behavior*, *25*(4), 887–896. https://doi.org/10.1016/j.chb.2009.03.003 (accessed 30 December 2023).

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, *76*(2), 149–188. https://doi.org/10.1006/obhd.1998.2804 (accessed 28 December 2023).

Meyerowitz, B. E., & Chaiken, S. (1987). The effect of message framing on breast self-examination attitudes, intentions, and behavior. *Journal of Personality and Social Psychology*, *52*(3), 500–510.

Müller, C., Graaf, B. & Pfeiffer, K. (2022). *World Robotics 2022 – Service Robots: Executive Summary*. Frankfurt a.M., Germany. IFR Statistical Department, VDMA Services GmbH. https://ifr.org (accessed 5 January 2023).

Naneva, S., Sarda Gou, M., Webb, T. L., & Prescott, T. J. (2020). A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. *International Journal of Social Robotics*, *12*(6), 1179–1201. https://doi.org/10.1007/s12369-020-00659-4 (accessed 26 December 2023).

Phillips, E., Zhao, X., Ullman, D., & Malle, B. F. (2018). What is human-like? In: T. Kanda, S. Ŝabanović, G. Hoffman, & A. Tapus (Eds.), *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 105–113). ACM. https://doi.org/10.1145/3171221.3171268 (accessed 27 December 2023).

Roth, S., & Nazemian, N. (2019). Brand Citizenship wertebasiert, generationsgemäß und pragmatisch aufbauen. In: F.-R. Esch, T. Tomczak, J. Kernstock, T. Langner, & J. Redler (Eds.), *Corporate Brand Management* (pp. 357–379). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-24900-7_19 (accessed 3 January 2023).

Sarrica, M., Brondi, S., & Fortunati, L. (2020). How many facets does a "social robot" have? A review of scientific and popular definitions online. *Information Technology & People*, *33*(1), 1–21. https://doi.org/10.1108/ITP-04-2018-0203 (accessed 30 December 2023).

Scholz, C. (2014). *Generation Z: Wie sie tickt, was sie verändert und warum sie uns alle ansteckt* (1st ed.). Wiley-VCH. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1101398 (accessed 27 December 2022).

Solomon, M. R. (2017). *Consumer Behavior [electronic resource]: Buying, Having, and Being*. Pearson.

Stapels, J. G., & Eyssel, F. (2021). Einstellungen gegenüber sozialen Robotern. In: O. Bendel (Eds.), Business book summary. *Soziale Roboter: Technikwissenschaftliche, wirtschaftswissenschaftliche, philosophische, psychologische und soziologische Grundlagen* (pp. 231–250). Springer Gabler.

Thaler, R. H. (2021). *Nudge: The Final Edition* (Updated edition). Penguin.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453–458. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1101398 (accessed 2 January 2023).

Irina Simon

Chapter 13
Towards a Structuralist Data Narratology

Abstract: A data story is a form of data-driven visualisation, that utilises narrative elements to turn data into actionable insight. This chapter presents the results of an interdisciplinary structuralist analysis of a corpus of 100 business data stories from the two data story galleries: Tableau Public and Power BI. In the analysis, three distinct components were identified, which created the basis for three recurring economic narrative patterns. In light of the growing deployment of artificial intelligence (AI) solutions in business, the research question driving the analysis was: What are the predominant business data storyline types and how can AI assist in the processes of authoring and typologising data narratives? Prior research has not identified storylines in business data narratives. This chapter aims at filling the missing gap in literature on narrative patterns in business data stories and seeks to inform a structuralist business data narrative method.

Keywords: data narrative, data story, artificial intelligence, narratology, storytelling, data visualisation, literary theory, narrative economics, computational narratology

13.1 Introduction

Stories mirror economies in contents and format, because, directly or indirectly, they thematise economic developments and utilise repeating patterns in structure, which is economical for both author and reader.

Boccaccio's *Decameron* (Boccaccio, 2007 [1886]) – a collection of 100 novellas written during the Black Death – is an early renowned example. In *Grammaire du Décaméron,* Todorov (1969b) detects repeating narrative patterns as all actions can be summed up to variations of either: modification, transgression or punishment (Tyson, 2006). For Todorov, the three actions are symptomatic of society's demand for an end to the medieval system of exchange based on feudalism and barter. Italy was on the cusp of Renaissance which gave way to a more liberal and *monetary* economic order. In it, traders, bankers and merchants established a novel system of commerce: capitalism. Thus, he writes:

> Si le livre a un sens général, c'est bien celui d'une libération dans l'échange, d'une rupture dans l'ancien système au nom de l'audacieuse initiative personnelle. [. . .] L'idéologie de la nouvelle

Irina Simon, Ruhr-West University of Applied Sciences, Mülheim an der Ruhr,
e-mail: irina.simon@hs-ruhrwest.de

https://doi.org/10.1515/9783110790320-013

> bourgeoisie consiste précisément à contester l'ancien système de l'échange, devenu trop étroit, et, à sa place, d'imposer un autre, plus 'liberal', qui peut faire croire, au début surtout, qu'il consiste en la disparition totale de système. (Todorov, 1969b, p. 81)

Translated and paraphrased into English by the author, Todorov claims that the main theme of *The Decameron* is that of freeing society of the constraints of the old system of exchange in the name of bold personal initiative. The renewal of systems is not just a message in terms of contents but also in form as the stories are told in narrative prose fiction in Florentine vernacular.

Following this approach, this chapter presents the results of an interdisciplinary structuralist analysis of a corpus of 100 business data stories. In the analysis of these data narratives, three distinctive features were identified, which created the basis for three economic storylines. Prior research on data stories has focussed on narrative patterns in data journalism (Matei & Hunter, 2021; Ojo & Heravi, 2018; Segel & Heer, 2010), however, business data storylines have not been identified. Research has proven the positive effects of data storytelling competency on business performance and decision-making (Daradkeh, 2021), and this chapter specifies the role of narrative patterns in this regard. Thus, the research question driving this analysis is: What are the predominant business data storyline types and how can artificial intelligence assist the processes of authoring and typologising business data stories?

13.2 What Is a Data Story?

An analysis of data narratives and the potential development of a data narratology[1] requires a definition of data stories. As a key component of data narratives, a definition of *narrative* (or story) comes first. There is a shared understanding over the fact that, storytelling is a form of sense-making; it provides explanations, and thereby enhances teaching and learning (Matei & Hunter, 2021; Shiller, 2019). The moment a narrative manages to inspire a change in thinking, it becomes tellable (Matei & Hunter, 2021; Ryan, 1986).

In contrast, there is no consensus about how to define and distinguish data narratives from other forms of data visualisation. In the following, a few academic definitions will be presented and contrasted, leading to two definitions: a wider sense of the word and a narrower one.

1 The term *narratology* was coined by Todorov as the study of narratives and universal narrative patterns. Such studies are structural, as they do not involve individual stories, characters, metaphors and language (sjuzhet), but are directed towards larger and recurring semantic patterns of events in (chrono-)logical order (fabula). A simple form of a universal plot pattern, may be the transition from one equilibrium to a new one. The two states of stability are separated by an imbalance, which can mark a conflict (or violation of an order) which is restored in the new equilibrium (Todorov, 1969a).

13.2.1 Operational

A common way to define data stories (and synonymously data narratives) is by describing what their "actions" are. An oft-cited description of data storytelling is that *(a) data (story) tells a story (with data)* (Borges, Correa & Silveira, 2022; Segel & Heer, 2010; Sekar, 2022). While this data-oriented operational definition focuses on agency, it promotes data or an entire data story to narrator. Its use of personification and tautology renders this explanation futile. Focussing on actions and purpose without personification and tautology, Mauro (2021) defines a data story as a tool for transforming data into business action. Similarly, Eckert (2022) defines one kind of data story as serving the purpose of providing an explanation to a connection. Yet, a second key form of data storytelling as found in data journalism contains data as primary subject of a story (Eckert, 2022).

13.2.2 Compositional

When data becomes the subject of a story or when the story is *about* data, it is defined by composition. Visual data stories may be defined by the various story elements they contain (e.g., Lee et al., 2015; Ramm et al., 2021; Segel & Heer, 2010; Shi et al., 2021; Sun et al., 2022), such as context, visuals, simplification, focus and coherence (Sekar, 2022).

13.2.3 Examples

As prime example in his definition, Sekar (2022) analyses a multi-page dashboard – a form of data visualisation showing numerous Key Performance Indices (KPI) at one glance. In contrast, multiple authors define dashboards as *opposed to* or *part of* data stories and not synonymously (Lee et al., 2015; Salleh, Mohamed & Shah, 2021; Serhal 2021; Tableau Help Stories). This suggests that an explanation based on an example does not suffice, as it may lead to the inclusion or exclusion of dashboards.

With these considerations in mind, we arrive at the following definitions:

Data story in the narrow sense of the word is a visual, interactive narrative with several views or pages containing data sets with context and development, serving the purpose of teaching and learning.

Data story in the broader sense is a form of data-driven visualisation regardless of extent that utilises any narrative element to turn data into actionable insight.

As prime topic of this chapter, business data stories denote either narrow or broad-sense narrative visualisations with economy-related data, addressing internal or external stakeholders with informative intent.

13.3 Corpus and Method

The corpus for this chapter contains only data stories in the narrower sense which are longer than just one page or single-view dashboard. The corresponding corpus of 100 data stories consists of 36 "Business Dashboards" from the online platform Tableau Public (Tableau Public)[2] and 64 data stories from the "Data Stories Gallery" by Power BI (Power BI) in the category "Business" filtered by "Top Kudos."[3] Both online galleries showcase best practices in data storytelling and visualisation and provide a platform for community building.[4]

In the documentation of the findings, each example was listed with metadata: name, date and author. As first step into the analysis, each entry of the corpus has been attributed with analytical data: page number, summary, level of interaction and narrative features.

The basic analytic procedure for the systematisation of narratives was derived from the narratological approach by Todorov, in which each story is searched for structural elements and their function in light of the overall plot. The structural elements are then developed into narrative. In combination, Todorov's decisive plot actions (punish, transgress and modify) form a common storyline of equilibrium, disequilibrium and new equilibrium (Todorov, 1992).[5] In the following, Todorov's plot actions and storyline combinations are correlated with the types derived from the qualitative content analysis of the data story corpus. Todorov's structuralist approach is a cross-disciplinary precursor to mixed qualitative content analysis involving steps analogous to inductive category formation and type-building content analysis (Mayring, 2014, p. 79 and p. 105). More specifically, after summarising the main points of each data story, categories were defined with dimensions and criteria. The initial inductively developed

2 The data stories from Tableau Public amount to 36 multi-page or multi-view data narratives retrieved on 30 January 2022. Of the 103 available data stories, all were accessible and operational. The dashboards on Tableau serve as business accelerators or demos and often use "real world/fake data" (RWFD) – a community project to create real-world dashboards with generated data.

3 The 64 data stories from Power BI's community "Data Stories Gallery" include all available and functional dashboards with multiple views from the first 10 webpages with 16 data narratives each. The 64 data stories in the corpus were derived by excluding all non-functional and single-view data stories from the set of 160. The Power BI sample also contains generated data, for example the GitHub United Nations Volunteers Challenge. The author acknowledges possible limitations associated with the data set, as some of the stories were built upon generated data and may contain predesigned story elements, yet highly applicable and top ranked on the two platforms.

4 Tableau Public, which is the self-labelled "largest repository of data stories," offers its users an interactive forum to share and feedback on data visualisations and to connect to the Tableau community or "Data Fam." On Microsoft's Power BI platform users may also post, view, share and like content.

5 A variation of this story about recovered stability, "conversion" narratives may start in conflict (disequilibrium) and then establish new balance throughout the sequence of events (Todorov 1992).

category measures narrativity in levels of completion of the tripartite narrative pattern: from full story (a, b, c), to short narrative sequence (a, b), to story fragment (a).

13.4 Business Data Story Typology

Data narratives can be systematised according to various parameters in many disciplines (e.g., data journalism,[6] literary criticism,[7] interdisciplinary[8]). A uniform framework for narrative typology has not yet been agreed upon, neither for data stories in general, let alone business data stories. In the following, the prototypical types derived from inductive category formation are laid out as main business storylines. Each storyline stands for a type of data story with specific economic purpose indicated by its corresponding narrative elements. The three types form a dialectical thesis-antithesis-synthesis triad.

13.4.1 Compliance – Expansion

This is the story of a business operation that runs according to plan along the value chain. In it, a transaction is planned, organised, staffed, lead, controlled, remunerated and evaluated according to protocol. While this story may have limited tellability (Ryan, 2005) or newsworthiness, it is a prominent *fabula* in business communication. This plot is about a plan successfully put into action and ultimately *compliance*. The narrative value of this story for businesses lies with the duplication of one optimal performance which is realised in mass production. The reproduction of a successful business plan or process in an organisation, presupposes attention to *detail*.

6 In their remarkable research on narratives in data journalism, Segel and Heer (2010) identify seven genres, narrative structures, approaches (reader-driven and author-driven) and visualisation structures. Ojo and Heravi (2018) also distinguish seven types of data stories in data journalism, which either disprove information or provide a specific kind of information.
7 Literary criticism identifies multiple story patterns. Most notably, they revolve around the main character, or hero, his journey and psychological effect on them (Campbell, 1993 [1949]). The number and scope of plot patterns vary from one universal monomyth (Campbell, 1993 [1949]) to a compendium of 36 storylines (Polti, 1921 [1916]).
8 In their interdisciplinary approach, Dillon and Craig (2022) classify four types of recurring stories to be found across many disciplines in both fictional and non-fictional texts: modelling, point of view (framing), identities and anticipation.

13.4.2 Rationalisation – Reduction

Opposite to the enabling compliance story with its project-oriented mindset, lies a problem-oriented one. This story is about eradicating flaws, as promoted by lean management theory and others. In it, a business function is *not*, or seemingly not, performing as planned. Business settings with this story may include, but are not limited to: using up too many resources (expenses, manpower, time, etc.), creating too much excess wastes or impaired functionality. In order to find areas for improvement, strategy must involve *critical* reflection with comparison and contrast.[9]

13.4.3 Transformation – Change

The third story contains a synthesis of compliance and rationalisation. It involves a fundamental change, because new developments necessitate modification. Business executives initiating organisational transformations face strategic decisions that have to be weighed against potential benefits and risks. The transformation or change story in business, involves a consideration of alternatives and potential strategies using interaction, simulation or scenario techniques.[10]

Table 13.1 shows the three stories as prime representatives for the types, each with their inductive categories in juxtaposition.

13.5 Distribution

The distribution of narrative elements and consecutive patterns in the corpus is shown in Table 13.2 and Figure 13.1. Detail is the defining component for data storytelling and was the prevailing narrative element found in the entire corpus.

Detail as exclusive narrative element, without pairing, occurred in 35% of all data stories. This makes the compliance story the second most common metanarrative in data storytelling. Contrast and comparison in combination with detail is the most common narrative, occurring in more than half of the data stories. The least frequent

9 While a detailed view on the key performance indices may already contain information that can be contrasted and compared, it is often primarily designed to bolster data collation. For example, when several sub-categories are displayed with individual detail on separate pages of the dashboard or with a different type of chart, the focus is more on illustrating their individual strengths than on their relative specifics.

10 The respective story pattern corresponds with the level and quality of data preparation or the analytical depth of the data set. While some data sets are already pre-structured, for example divided into broad categories or even arranged into groups with commonalities, other sets may see complex interrelated systematisation, or include trends and recommend multiple actions with respective outcome.

Table 13.1: Prime representatives for the types found in corpus in form of storylines with inductive categories. (compiled by author.)

	Compliance	**Rationalisation**	**Transformation**
Strategy	Abidance, repetition, reproduction	Fault detection, reduction, deviation	Modification, change, adaptation
Method	Intensification, specialisation, particularisation	Critical inquiry, juxtaposition, outlier	Context, enlarge portfolio and perspective
Interaction	Some	Little	High
objective	Mass production	Streamlining, termination	Adjustment, transition
Principle in economics	Economies of scale	Diseconomies of scale	Economies of scope
movement	Forward	Backward, stand-still	Anywhere, new direction
Mode (see Figure 13.2)	Descriptive, prescriptive	Descriptive, explorative	Predictive, substantive
Data narrative element	Detail (a)	Compare & contrast view (a, b)	Interactive simulation, scenario (a, b, c)
Todorovian storyline	Equilibrium (punish)	Equilibrium + disequilibrium (punish + transgress)	Equilibrium + disequilibrium + new equilibrium (punish + transgress + modify)

Table 13.2: Distribution of narrative types among corpus and platforms. (compiled by author.)

	Compliance detail (uncombined)	**Rationalisation detail + contrast**	**Transformation detail + contrast + simulation**
Power BI (total 64)	19 (30%)	33 (52%)	12 (19%)
Tableau (total 36)	16 (44%)	19 (53%)	1 (3%)
Overall (total 100)	35 (35%)	52 (52%)	13 (13%)

narrative pattern was transformation with an occurrence of 13% and a particularly higher incidence in the Power BI gallery (19%), marking the change story a less salient one in the context of business data stories.

13.6 Utilising AI for a Data Narratology

There is no evidence for research on AI-supported plot analysis and data story categorisation, owing to the fact that data stories are multi-media representations using visualisation *and* narrative.

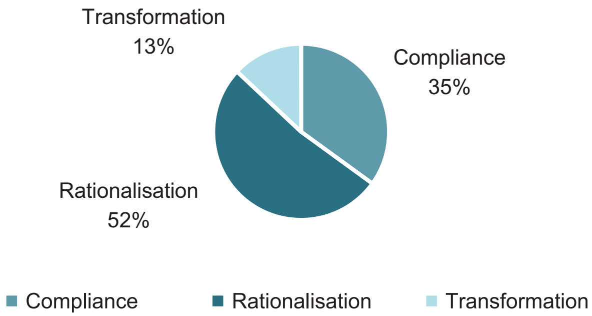


Figure 13.1: Distribution of metanarratives. (compiled by author.)

For example, computational narratology analyses how AI systems can perform human-like narrative (Mani, 2014). In it, algorithms are trained with the help of NLP to parse various literary narrative patterns of, for example Vladimir Propp, Joseph Campbell, Gustav Freytag and Gérard Genette in annotated literary corpora (Mani, 2014). Recent advances in machine learning (ML) have made complex and large-scale corpora mining possible (Authors.ai, 2021). Similarly, Sun et al. (2022) list various articles and programs related to the area of automatic data visualisation, including research on rule-based and ML-based automatic data visualisation and analysis. Neither of these approaches specifically investigates the characteristics of *data narratology,* as *both* a visual and narratological object of study.

Despite the enormous progress in the field of text generation, Segel and Heer (2010) declare, still topically, that the data visualisation tools on the market mainly explore and analyse data, but offer very limited support to develop visual narratives (Obie et al., 2019). While AI may optimise a literary narrative (see e.g., the self-editing AI Marlowe, Authors.ai, 2021) or enhance a visualisation (e.g., by providing a short description to a visualisation, see Tableau Data Story or the Smart Narrative Function in Power BI), its capabilities to generate new inspiring insight by *developing* and *combining* narrative and visuals, are still lacking. Tableau offers its users the feature “Data Stories” which turns any given data set into visualisation and text with the help of “rule-based templated natural language generation (NLG)” (Tableau, Data Stories). Still, Tableau notes that “Data doesn’t use generative AI, large language models (LLMs), or ML to write insights or stories” (Tableau, Data Stories).

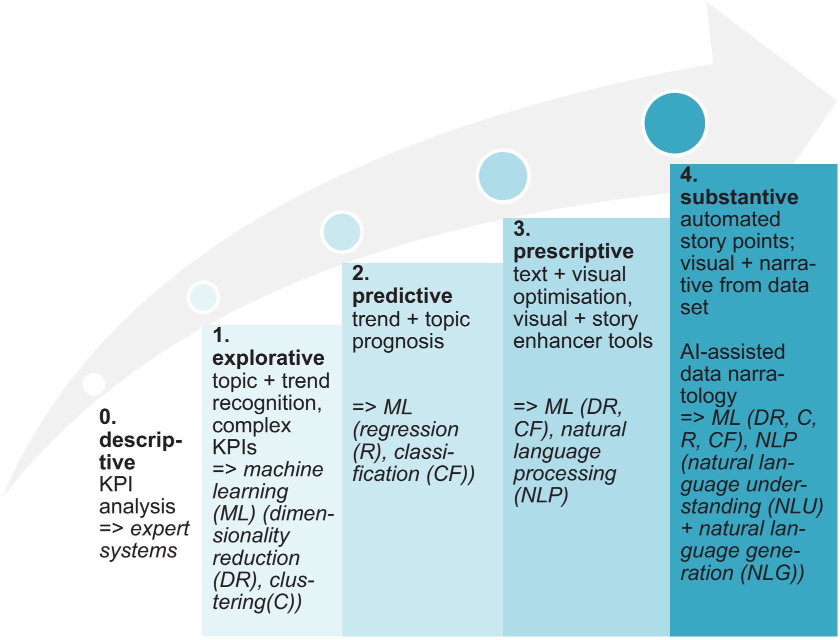


Figure 13.2: AI-utilisation for business data narratology and data storytelling support. (translated, adapted and modified from Eckert, 2021; Veel, 2018; Badillo et al., 2020).

Concerning AI-deployment in storytelling, Eckert identifies five levels of AI-utilisation in storytelling. These stages were modified and specified for data stories by the author, as shown in Figure 13.2.

Starting at rule-based analytics of KPIs with a descriptive formula, technically not involving AI, the figure shows a steady increase in AI maturity and autonomy with each consecutive stage and task complexity. ML may be deployed at the second level, recognising patterns in topics and trends. Predictive practices take place in the second stage, followed by prescriptive text-optimisation tools. The final stage, with the highest level of autonomy and the most complex tasks, is labelled *substantive* in this figure, as it eventually generates suitable story points from data to be implemented in business data narratives. The final two stages, prescriptive and substantive, are central to the two key objectives of the research question: (a) defining the dominant data narrative patterns as significant step towards a data narratology with distinctive storyline types and (b) improving the process of authoring data stories with the help of AI. Both Power BI and Tableau do not support fully automated data story generation from data sets and are also not yet capable of typologising business data narratives. They do, however, offer assistance in the authoring process of data story *points*, transforming data sets into visualisation and providing descriptions to charts, equivalent to mostly descriptive and limited substantive AI-utilisation practices (Tableau Help "Choose the Right Story Type" and Microsoft Power BI "What is data storytelling?").

13.7 Interpretation

Assuming that the economic metanarrative is about social provisioning and the creation, increase, and preservation of value(s), the prevalence of the second storyline does not come as a surprise. It is the reductionist narrative that best resembles the sustainability efforts of the global economy in the 21st century, as it revolves around conscious consumption and optimal use of economic means and, eventually, achieving the same productivity with *fewer* resources, especially non-renewable ones.

In light of the findings from the analysis, the benefits of AI-utilisation are controversial, as AI draws from existing data and will most likely reproduce common patterns. Thus, the use of AI to improve data storytelling at this point, comes with the threat of authoring more predictable (see Mindner, Schlippe & Schaaff, 2023, p. 6), and ultimately less tellable, stories. Accordingly, Veel (2018) points out that NLG may "decode" data into narrative or discover "hidden" stories, but never "make" meaning like humans do. For example, Veel (2018) describes how natural language understanding (NLU) could structure and disambiguate a large pool of data stories and, after that, NLG could summarise and generate metanarratives from the structured data. As stated before, the gain of this strategy is debatable since existing samples lack originality. Moreover, from an ethical point of view, one has to ask "[. . .] if narrative is regarded as an intrinsically human form of meaning-making, then what are the implications of automating this process?" (Veel, 2018, p. 4)

Efforts to change the trend of pure pattern recognition and compliance in storytelling by promoting variety and pluralism of perspectives in stories may involve intelligence augmentation (IA) (Harborth & Kümpers, 2022), as opposed to, or in combination with, artificial intelligence. IA includes augmented reality (AR) and virtual reality (VR). Both AR and VR are highly immersive, which can substantially improve user engagement and the immediacy of data narrative as well. A higher level of media richness and interaction can improve involvement and thereby deliver more varied business insight.

Either way, in order to increase immersiveness, data narratives need to become more reader driven. Regarding the corpus, the predominant use of linear progress, little interaction and pre-determined message do not enable the reader. With the reader empowered to co-author of the story, in line with the process of "reading as construction" (Todorov, 1980), data stories could inspire and create lasting impact, which may drive better decision-making and innovation. According to this argument, narrative "[. . .] should be regarded as an intertextual patchwork drawing upon multiple sources, which coalesce in the reader" (Veel, 2018, p. 6). With this idea of more diversification in concert with reader empowerment, each pivotal point would become an interactive narrative junction, offering readers more than one alternative to create and rewrite the data story and its path to their own liking. Of course, various internal and external business environments and differing intents may create greater or lesser needs for variety and reader empowerment.

13.8 Outlook

There is evidence to believe that data stories themselves may be on the brink of transition. According to Serhal, data storytelling was introduced to the market in order to support dashboards (the most popular tool for the last three decades), but could never replace them. This is due to the fact that data stories and infographics – as forms of storytelling for business users – are single-use, whereas monitoring dashboards can be used multiple times. At this point, Serhal introduces *industrialised storytelling* which has recurring utility for business professionals. This form of narration comes with built-in narrative techniques, including comment, glossary and call to action. Industrialised storytelling promises more efficiency and effectivity for data storytelling.

Concluding, it became apparent, that further exploration of narrative in its various forms, including data narratives and potentially industrialised storytelling, is necessary. The predominant metanarratives in the corpus were: compliance, rationalisation and transformation, with the transformative change story occurring with least frequency. Since all of these narratives are recurring patterns, AI and ML can be employed to improve the writing process of data stories, but the little variation in narrative patterns and innovative deviation in form and style will continue to necessitate human creativity (Zhang, et al., 2022). AI-driven storytelling is highly dependent on and determined by the data utilised in the training of its algorithms. Thus, its orientation is retrospective, and operates in the tradition of pre-existing texts and visuals, but cannot (yet) generate versatile meaning and authentic original thought.

As an outlook, these steps are proposed in order to facilitate further research in data narratology and continued development of data story tools:

- advance analysis of data stories to develop a data narratology (e.g., by applying computational narratological frameworks onto data stories and other forms of multi-media narratology)
- promote more variety in data narratives (e.g., by endorsing innovative data story authors and their communities)
- lower the barriers to AI-adoption and research with AI (e.g., by funding projects in data narratology and testing grounds).

Although data storytelling, as a medium, is a more recent form of data visualisation, the analysis of the corpus did not reveal a collective wish for change. So, in analogy to

The Decameron, the format can be considered progressive, but unlike Bocaccio's collection of stories, there is no evidence for a new era in commerce or a growing interest in the transformative story in business data narratives. Still, analysing data stories is, and will continue to be, vital as data stories contain individual and collective economic insight (Dillon & Craig, 2022 referencing Shiller, 2019). Ultimately, internal and external business stakeholders need to become data storylisteners (Dillon & Craig, 2022) and data narratologists as an extended understanding of data narratives and their composition can, and will, inform better decision-making and enable economic agents to reimagine the future.

Bibliography

Authors.ai. (2021). Marlowe Analyzes *The Da Vince Code*. https://authors.ai/ (accessed 30 January 2023).

Badillo, S., Banfai, B., Birzele, F., Davydov, I.I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., & Zhang, J.D. (2020). An introduction to machine learning. *Clinical Pharmacology and Therapeutics*, *107*(4), 871–885. https://doi.org/10.1002/cpt.1796.

Boccaccio, G. (2007) [1886]. *The Decameron of Giovanni Boccaccio*. https://www.gutenberg.org/ebooks/23700 (accessed 4 December 2022).

Borges, M., Correa, C.M., & Silveira, M.S. (2022). Fundamental elements and characteristics for telling stories using data. *Journal on Interactive Systems*, *13*(1), 77–86. https://doi.org/10.5753/jis.2022.2330.

Campbell, J. (1993) [1949]. *The Hero with a Thousand Faces*. London: Fontana Press.

Daradkeh, M.K. (2021). An empirical examination of the relationship between data storytelling competency and business performance: The mediating role of decision-making quality. *Journal of Organizational and End User Computing*, *33*(5), 42–73. http://doi.org/10.4018/JOEUC.20210901.oa3.

Dillon, S., & Craig, C. (2022). Storylistening: A case study in how to include the humanities in evidence provided for public reasoning. *Journal of the British Academy*, *10*, 21–28. https://doi.org/10.5871/jba/010.021.

Eckert, H.-W. (2021). Kann Künstliche Intelligenz Stories Schreiben? https://www.marconomy.de/kann-kuenstliche-intelligenz-stories-schreiben-a-1035723/ (accessed 30 January 2023).

Eckert, H.-W. (2022). *Storytelling with Data: Gaining Insights, Developing Strategy and Taking Corporate Communications to a New Level*. Wiesbaden: Springer Gabler.

Gottfredson, L.S. (1997). Mainstream science on intelligence. *Intelligence*, *24*(1), 13–23. https://doi.org/10.1016/s0160-2896(97)90011-8.

Harborth, D., & Kümpers, K. (2022). Intelligence augmentation: Rethinking the future of work by leveraging human performance and abilities. *Virtual Reality*, *26*, 849–870. https://doi.org/10.1007/s10055-021-00590-7.

Knaflic, C.N. (2015). *Storytelling with Data*. Hoboken, NJ: John Wiley & Sons.

Lee, B., Riche, N.H., Isenberg, P., & Carpendale, S., (2015). More than telling a story: Transforming data into visually shared stories. *IEEE Computer Graphics and Applications*, *35*(5), 84–90. https://doi.org/10.1109/MCG.2015.99.

Mani, I. (2014) Computational narratology. In: P. Hühn, J. C. Meister, J. Pier and W. Schmid (Eds.), *Handbook of Narratology*, 84–90. Berlin/Boston: DeGruyter.

Matei, S.A., & Hunter, L. (2021). Data storytelling is not storytelling with data: A framework for storytelling in science communication and data journalism. *The Information Society*, *37*(5), 312–322. doi: 10.1080/01972243.2021.1951415.

Mauro, A.D. (2021). *Data Analytics Made Easy: Analyze and Present Data to Make Informed Decisions Without Writing Any Code*. Birmingham/Mumbai: Packt Publishing.
Mayring, P. (2014). *Qualitative Content Analysis: Theoretical Foundation, Basic Procedures and Software Solution*. Klagenfurt. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-395173 (accessed 30 January 2023).
Microsoft Power BI Community. Data Stories Gallery: Sort by: Top Kudos in Category: Business. https://community.powerbi.com/t5/Data-Stories-Gallery/bd-p/DataStoriesGallery?sortby=kudos&filter=business (accessed 30 January 2023).
Microsoft Power BI. What is data storytelling? https://powerbi.microsoft.com/en-au/data-storytelling/ (accessed 30 January 2023).
Mindner, L., Schlippe, T., & Schaaff, K. (2023). Classification of Human- and AI-Generated Texts: Investigating Features of ChatGPT. https://doi.org/10.48550/arXiv.2308.05341.
Obie, H.O., Chua, C., Avazpour I., Abdelrazek, M., Grundy, J., & Bednarz, T.A. (2019) Framework for authoring logically ordered visual data stories. *IEEE Symposium on Visual Languages and Human-Centric Computing* (VL/HCC), 257–259. https://doi.org/10.1109/VLHCC.2019.8818925.
Ojo, A., & Heravi, B. (2018). Patterns in award winning data storytelling. *Digital Journalism*, *6*(6), 693–718. https://doi.org/10.1080/21670811.2017.1403291.
Piper, A., So, R J., & Bamman, D. (2021) Narrative Theory for Computational Narrative Understanding. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 298–311.
Polti, G. (1921) [1916]. *The Thirty-Six Dramatic Situations*. Franklin, Ohio: James Knapp Reeve.
Ramm, S., Kopf, E.-M., Dinter, B., & Hönigsberg, S. (2021) What Makes a Good Story – The Use and Acceptance of Storytelling in Business Intelligence. *Proceedings of the 54th Hawaii International Conference on System Sciences*, 5706–5715. http://hdl.handle.net/10125/71312.
Ryan, M.L. (1986). Embedded narratives and tellability. *Narrative Poetics*, *20*(3), 319–340.
Ryan, M.L. (2005). Tellability. In: D. Herman et al. (eds.) *Routledge Encyclopedia of Narrative Theory*. (pp. 589–91). London: Routledge.
Salleh, A.A., Mohamed, N.A., & Shah, N.A.S.A. (2021) Simulating data stories of clients' credit card default. *Applications of Modelling and Simulation*, *5*, 184–190.
Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1139–1148. https://doi.org/10.1109/TVCG.2010.179.
Sekar, M. (2022). *Machine Learning for Auditors*. Berkeley, CA: Apress.
Serhal, J.P. (2021). Industrialized Storytelling: Positioning & Use Cases. https://www.toucantoco.com/en/blog/industrialized-storytelling-positioning-use-cases (accessed 30 January 2023).
Shi, D., Xu, X., Sun, F., Shi, Y., & Cao, N. (2021). Calliope: Automatic visual data story generation from a spreadsheet. *IEEE Transactions on Visualization and Computer Graphics*, *27*(2), 453–463.
Shiller, R.J. (2019). *Narrative Economics: How Stories Go Viral & Drive Major Economic Events*. Princeton and Oxford: Princeton University Press.
Sun, M., Cai, L., Cui, W., Wu, Y., Shi, Y., & Cao, N. (2022). Erato: Cooperative data story editing via fact interpolation. *IEEE Transactions on Visualization and Computer Graphics*, *29*(1), 983–993. https://doi.org/10.1109/tvcg.2022.3209428.
Tableau Public. Business Dashboards. https://public.tableau.com/app/discover/business-dashboards (accessed 30 January 2023).
Tableau Help. Stories. https://help.tableau.com/current/pro/desktop/en-gb/stories.htm (accessed 30 January 2023).
Tableau Help. Choose the Right Story Type for Your Tableau Data Story. https://help.tableau.com/current/pro/desktop/en-us/data_stories_story_types.htm (accessed 30 January 2023).
Tableau Help. Create a Tableau Data Story (English Only). https://help.tableau.com/current/pro/desktop/en-gb/data_stories.htm (accessed 30 January 2023).
Todorov, T. (1969a). Structural analysis of narrative. *Novel: A Forum on Fiction*, *3*(1), 70–76.

Todorov, T. (1969b). *Grammaire du Décaméron*. The Hague/Paris: Mouton.
Todorov, T. (1980). Reading as construction. In: S. R. Suleiman & I. Crosman (Ed.), *The Reader in the Text: Essays on Audience and Interpretation* (pp. 67–82). Princeton: Princeton University Press. https://doi.org/10.1515/9781400857111.67.
Todorov, T. (1992) [1977]. *The Poetics of Prose*. Ithaka, New York: Cornell University Press.
Tyson, L. (2006). *Critical Theory Today: A User-Friendly Guide* (2nd ed.). New York: Routledge.
Veel, K. (2018). Make data sing: The automation of storytelling. *Big Data & Society*, *5*(1). https://doi.org/10.1177/2053951718756686.
Zhang, Y., Reynolds, M., Lugmayr, A., Damjanov, K., & Hassan, G.M. (2022). A visual data storytelling framework. *Informatics*, 2022, *9*(73), 1–22. https://doi.org/10.3390/informatics9040073.

Khuliso Mapila, Tankiso Moloi

Chapter 14
Exploring the Adoption of AI for Customer Engagement Marketing by Small and Medium Enterprises in South Africa: A Literature Review of Challenges and Opportunities

Abstract: In South Africa, small, micro, medium enterprises (SMMEs) are vital. However, majority fail in the first few years from inception due to various factors, more often marketing-related problems. One of the fastest-evolving technologies, artificial intelligence (AI), has enormous potential in marketing. South Africa has low AI adoption rates which limit small businesses' growth, competitiveness and sustainability. This research provides a literature review on the adoption challenges and opportunities within SMMEs for customer engagement marketing. This article conducts a literature review and content analysis. Scopus and Google Scholar are used to identify relevant articles on AI adoption and marketing. The report summarised research quality, data and search results. The results suggest various challenges and opportunities in AI adoption. This research is conceptual, so future studies should validate the proposed challenges and opportunities for South African SMMEs. This research is one of the newest to examine the adoption of AI in marketing by South African SMMEs.

Keywords: artificial intelligence, marketing, literature review, customer engagement marketing, AI adoption, SMMEs

14.1 Introduction and Background

Small, micro, medium enterprises (SMMEs) play a critical role in fostering economic growth in South Africa by eradicating poverty and creating jobs (Bushe, 2019). By 2030, SMMEs are expected to create 90% of all new jobs, according to the National Development Plan (NDP) of the South African government, highlighting the significance of these small businesses (Bhorat et al., 2018). Given the high unemployment rate in South Africa, currently at 48%, SMMEs are especially important (Statistics South Africa, 2022). South Africa has a positive perception of entrepreneurial opportu-

Khuliso Mapila, Johannesburg Business School, University of Johannesburg, Johannesburg,
e-mails: 200908385@student.uj.ac.za; khully22@gmail.com
Tankiso Moloi, Johannesburg Business School, University of Johannesburg, Johannesburg,
e-mail: smoloi@uj.ac.za

https://doi.org/10.1515/9783110790320-014

nities, with SMMEs contributing 60% towards the South African labour force (Bhorat et al., 2018; Telukdarie et al., 2022).

According to SEDA's (2022) SMME Quarterly Update report for Q1, South Africa has over 2.4 million SMMEs, with 69% operating in the informal sector and 28% in the formal sector (SEDA, 2022). A large percentage of these businesses operate in the services sector which accounts for most of the South African output and jobs (Matekenya & Moyo, 2022). 30% of these businesses are in wholesale and retail trade, while 23% are in community and social services, and 14% of SMMEs are involved in financial services (Matekenya & Moyo, 2022). Although SMMEs are seen as important, they still face many challenges that can lead to business failure (Telukdarie et al., 2022). The ongoing fear of failure among existing and potential entrepreneurs (49.8%) juxtaposes the potential of creating established businesses in the country (Bhorat et al., 2018).

Businesses are now moving towards Industry 4.0, and technologies such as artificial intelligence (AI) are evolving. According to research on the reasons why various nations and businesses adopted AI, increased research and development spending, or created AI policies, many did so in order to compete with other rivals on the global market (Kabalisa & Altmann, 2021). However, the adoption of AI in most emerging markets and industries has been limited due to various constraints (Verma et al., 2021).

Organisations use AI in marketing to track real-time data, analyse it for better understanding, predict consumer needs and act on them quickly (Campbell et al., 2020; Verma et al., 2021). Businesses frequently utilise AI to discover new market opportunities and create marketing mix frameworks (Schlögl et al., 2019). A growing number of organisations are incorporating AI into their marketing strategies, which enables them to speed up process times and interact more directly with customers.

14.2 Research Problem and Objectives

14.2.1 Research Problem

However, AI is still in its growing phase, and in South Africa, the rate of technology adoption is slow, particularly among small businesses, which lag behind in AI adoption (Campbell et al., 2020). While some businesses want to include AI in every business decision, others struggle to navigate its adoption (Campbell et al., 2020).

Previous research has examined the benefits and drawbacks of using AI, with a focus on ethical and societal impacts, the impact on the workforce, and ways to improve decision-making and efficiency in various industries. However, despite a high potential AI has in marketing, limited research on the adoption of AI by SMMEs for marketing purposes and the opportunities and challenges of AI adoption in South Africa exist. This research aims to address the gap by examining the challenges and opportunities of AI adoption for customer engagement marketing in SMMEs. It will contribute to the understanding of AI adoption by SMMEs and the emerging literature on AI marketing for SMMEs (Alupo, Omeiza & Vernon, 2022; Huang & Rust, 2021; Matta et al., 2022; Regona et al., 2022; Taljaard & Gerber, 2022).

Based on the above, there is a need to understand the opportunities and challenges faced by SMMEs in order to potentially reduce the slow rate of AI adoption in South Africa. This can be accomplished by understanding the opportunities and challenges of adopting the technology within SMMEs.

14.2.2 Research Objectives

The potential impact of AI on organisations is expected to be significant, as it will shape the way firms interact with their consumers. (Hollebeek, Sprott, & Brady, 2021). It is on this basis that the purpose of this chapter is to compile a list of opportunities and obstacles for SMMEs to adopt AI. To this effect, the following research questions have been formulated:

- What are the existing challenges for AI adoption within SMMEs for customer engagement marketing?
- What are the existing opportunities for AI adoption within SMMEs for adoption of AI for customer engagement marketing?

The remaining sections in the chapter will present the following: literature review in Section 15.3; methodology in Section 15.4; results and findings in Section 15.5; and conclusion in Section 15.6.

14.3 Literature Review

14.3.1 Artificial Intelligence to Enhance Marketing

AI developments continue to expand globally, and South Africa leads Africa in AI start-ups and readiness (Arakpogun et al., 2021). South Africa has the most AI start-ups and the highest AI readiness ranking in Africa (Arakpogun et al., 2021).

AI is transforming how decision-makers in various fields, including marketing, make choices. Definitions of AI vary, with some considering it as “intelligence demonstrated by machines” (De Bruyn et al., 2020). Learning, observing, understanding, talking, interacting with others, planning, reasoning, creativity and problem-solving are among the tasks that AI can perform (Dimitrieska, Stankovska & Efremova, 2018). The capacity of AI applications, such as deep learning (DL), to generate high-order learning autonomously without relying on human intervention or specialised knowledge sets it apart from conventional statistical methods (De Bruyn et al., 2020).

AI processes data in many subdomains (Huang & Rust, 2021). Machine learning lets systems improve without programming. DL, a neural network that layers data and finds patterns, is another subfield (Campbell et al., 2020). Marketers should use three AIs: Mechanical AI automates repetitive tasks; Thinking AI analyses data to make decisions; and Feeling AI analyses and interacts with people (Huang & Rust, 2021).

Leaders value AI technologies that adapt to their business operations (Schlögl et al., 2019). AI can improve customer engagement by personalising experiences, anticipating needs and understanding sentiment. AI-based personalisation improves customer loyalty, sales and understanding. AI builds customer relationships through automated customer service and personalised recommendations. AI provides real-time customer behaviour insights for targeted marketing campaigns (Campbell et al., 2020; Dimitrieska, Stankovska & Efremova, 2018; Jarek & Mazurek, 2019; Verma et al., 2021).

It improves marketing research, strategy, planning, implementation, 4Ps (product, price, place, and promotion) management, targeting and positioning, customer response modelling, channels and logistics strategy, metrics and implementation control, and marketing data mining using artificial neural networks (Campbell et al., 2020; Haleem et al., 2022; Huang & Rust, 2021; Verma et al., 2021). AI can support finance, marketing and commerce decision-making, customer relationship support to boost loyalty, process acceleration to boost work efficiency, and knowledge management through data acquisition and processing (Schlögl et al., 2019). AI is used in commercial solutions for autonomous robots and vehicles, decision-making, voice recognition, image recognition and text recognition (Haleem et al., 2022; Jarek & Mazurek, 2019).

AI marketing's main benefit is personalisation and meaningful customer experiences (Haleem et al., 2022). AI can use customer data to personalise experiences. It can also target potential customers with relevant offers. AI can optimise campaigns in real time, improving marketing return on investment (ROI) (Campbell et al., 2020; Dimitrieska, Stankovska & Efremova, 2018; Haleem et al., 2022; Jarek & Mazurek, 2019; Verma et al., 2021).

With 5.3 billion mobile phone users and 4.95 billion internet users, modern technology and AI are growing worldwide (Deloitte, 2022; We Are Social, 2022). About 70% of South Africans use the internet and spend the most time (10:46 hours a day) on mobile phones, ranking fifth globally in terms of using the internet on mobile phones (We Are Social, 2022). Despite this high usage, only 17% of South Africans use voice assistance, 29% image recognition and 33% online translation tools to find information (Deloitte, 2022). The existence and current use of some applications allow South Africa to adopt modern technology, including AI (Dimitrieska, Stankovska & Efremova, 2018).

Businesses and marketers can expand by using AI marketing capabilities, but many are still unaware of the advantages or unsure of how to put AI to use to continue improving marketing (Campbell et al., 2020). At the moment, AI is subject to excessive exposure and experiencing rapid expansion, while business owners are still trying to fully grasp the difficulties and advantages of implementing the technology

(Schlögl et al., 2019). They are also still in the process of understanding the real impact that AI can have on their business and how to prepare themselves in order to adopt and use new technologies (Bettoni et al., 2021).

Although the opportunity exists, AI in marketing also has several challenges that need to be considered. Implementing AI technologies requires expertise, implementation and maintenance can be costly, and ethical issues must be considered (AlSheibani, Cheung & Messom, 2018; Dwivedi et al., 2021; Onyechi & Abeysinghe, 2009; Shaikh et al., 2021). The benefits to companies in using AI to enhance customer experiences exist (Dwivedi et al., 2021; Shaikh et al., 2021). The above is explored in detail, highlighting the opportunities and challenges in adopting AI.

14.3.2 Artificial Intelligence Adoption Theoretical Basis

There exist several adoption models and frameworks in evaluating the adoption of technology at organisational and individual levels. Literature shows that the most prominent AI adoption theories among the various adoption theories are the Technology, Organisation and Environment (TOE) and Diffusion of Innovation (DOI) which have both been used to research adoption both at individual and organisational levels (Afolayan & de la Harpe, 2020; Radhakrishnan & Chattopadhyay, 2020; Sithole & Ruhode, 2021). While both models aim to explain the factors that influence technology adoption, they differ in their focus and scope. TOE argues that various factors have been identified to be considered when adopting and implementing new technology, including technological factors, top management support – organisational factor and the size of the business (AlSheibani, Cheung & Messom, 2018). TOE is dominant within and discusses technology adoption at a business level, consisting of three contexts: Technological, Organisational and Environmental (Afolayan & de la Harpe, 2020; AlSheibani, Cheung & Messom, 2018). Unlike the DOI model, which primarily focuses on technological factors, the TOE model provides a more holistic view of technology adoption by considering a wide range of internal and external elements (Maroufkhani, Iranmanesh & Ghobakhloo, 2023).

Previous research demonstrated that the TOE framework is useful for examining adoption success factors for AI (Chen, Li & Chen, 2021). Because of its ability to cover all aspects of adoption issues and analyse the effect of organisational components on adopting new technology, the TOE model would need to be used in this chapter to examine the adoption challenges and opportunities in AI adoption for marketing by SMMEs in South Africa (Sithole & Ruhode, 2021). The research model, which is used to develop a hypothesis, includes the technological, organisational and environmental contexts for factors that pose challenges and opportunities for AI adoption for SMMEs. The model adapted by Sithole and Ruhode (2021) is highlighted below, as are the TOE factors influencing AI adoption to determine the driving opportunities and challenges that may be a barrier in adopting new technology applicable to SMMEs in South Africa. Stakeholders and small business owners can use the TOE framework to understand technological, organisational and environmental factors that affect AI adoption.

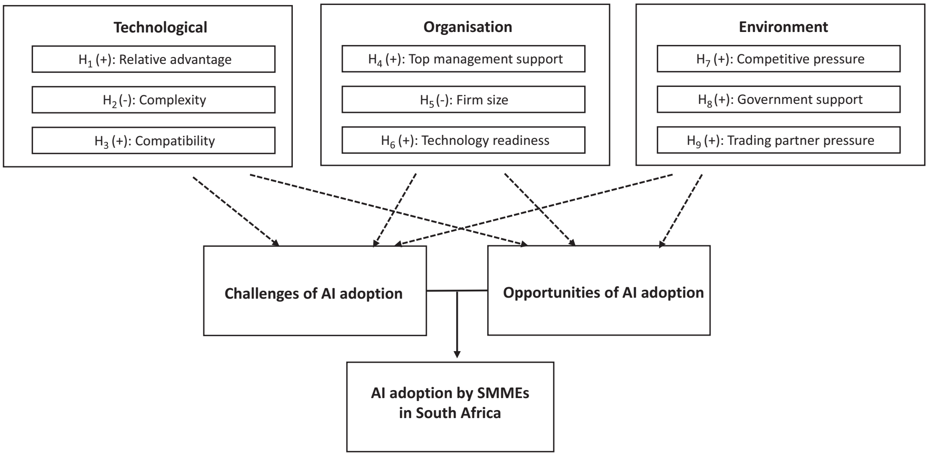


Figure 14.1: A conceptualised research framework is proposed for understanding the challenges and opportunities of AI adoption in marketing for SMMEs in South Africa (Sithole & Ruhode, 2021).

14.4 Methodology

The study was conducted following a literature-based review of the available literature on AI, marketing, customer engagement marketing and adoption. These articles were examined in light of the study's objective and research questions (Kofod-Petersen, 2012). This chapter conceptualises AI for customer engagement marketing challenges and opportunities. Scopus and Google Scholar were used to find these studies. The following search terms were taken from Dwivedi et al. (2021), which provided a comprehensive yet varied viewpoint on the various opportunities and challenges identified by numerous expert contributors. Industry, academia and government collaborators highlighted the significant opportunities, challenges and potential research agenda posed by AI in business and management, government and the public sector.

The following search keywords were used to gather results on Scopus and Google Scholar: (TITLE-ABS-KEY ((“Artificial Intelligence”) AND TITLE-ABS-KEY (“Advantages” OR “Benefit” OR “Opportunities” OR “Limitation” OR “Challenge” OR “Barriers” OR “Shortcoming”)) AND TITLE-ABS-KEY ((“Marketing”))). The selection contains keywords that are grouped synonymously into group sets to find all relevant literature (Kofod-Petersen, 2012).

14.5 Results and Discussion

14.5.1 Challenges and Opportunities of Adopting and Using AI

This search using the keywords above results in a total of 925 results on Scopus. The search was limited to papers, finalised articles and conference papers from 2000 to 2023. The inclusion of the keyword "Customer Engagement" returned only 54 results, indicating a limited focus on research relating to the customer engagement aspect of marketing and AI.

From a country perspective, the USA (110), China (93), India (75), UK (36) and Australia (36) publish the most on the topic. Few South African studies (3) returned studies using the above search terms and filters. Taking the above, the following section synthesises the opportunities and challenges of AI adoption based on previous literature consolidated through existing articles and past research.

Businesses face a new challenge and are presented with new opportunities as a result of the continued integration of technology into the world of work (Dimitrieska, Stankovska & Efremova, 2018). This highlights the importance of exploring the existing opportunities and challenges that small business owners face when adopting or using new technologies. The following sections propose a list of challenges and opportunities in adopting AI in marketing for small businesses.

14.5.2 Opportunities of Adopting AI for SMMEs

This section examines the advantages and opportunities of adopting AI for small businesses. AI is rapidly emerging as a powerful tool for businesses of all sizes. It has the potential to revolutionise how companies operate, allowing them to gain insights, automate processes and make better decisions. Small businesses, in particular, can benefit from AI by reducing costs, improving customer service, automatic processes, personalisation and customisation, improved marketing, chatbots, virtual assistants and increasing efficiency (De Mauro, Sestino & Bacconi, 2022; Matekenya & Moyo, 2022; Morgan, 2012; Ulrich & Frank, 2021; Venkatesh, 2022).

This literature review examines the advantages and opportunities of adopting AI for small businesses. The opportunities may vary by industry and sector; for example, in the construction industry, big data can reduce repetitive tasks, improve work processes, improve plan accuracy during design and planning, as well as producing results that stakeholders can recognise and improving stability and consistency because AI rarely makes mistakes (Regona et al., 2022). This is possible because AI has the capacity to automatically produce higher-degree constructs from raw data with little to no human involvement and to discover hidden patterns in data files (De Bruyn et al., 2020).

AI will enable process automation, freeing up marketers' time to focus on creativity. Additionally, it gives marketers new tools to attract, retain and satisfy customers. Additional benefits of AI may include smarter advertisements, improved content delivery and a greater reliance on Bots (Dimitrieska, Stankovska, & Efremova, 2018). AI's effects on marketers include the elimination of time-consuming and laborious tasks, the promotion of innovative design practices, new skills development within the marketing team, and the development of a new marketing ecosystem (Jarek & Mazurek, 2019).

Based on the literature review, the below highlights some opportunities found common among SMMEs from a global perspective in adopting AI technologies.

Competitive advantage: Technology, resources or strategic positioning can give a company a competitive advantage. A company that successfully adopts and uses AI technologies may be able to outperform its competitors by making faster and more accurate decisions, automating processes and gaining marketing insights from data. Competitive advantage in AI means a company can use AI to outperform its competitors, increasing market share, profitability and customer loyalty through personalised customer engagement (Dwivedi et al., 2021; Schlögl et al., 2019).

Economies of scale: A growing business can benefit from economies of scale. As a company invests in and adopts AI technologies, it may reduce costs and improve efficiency through automation and better decision-making. This can boost company profitability and competitiveness. Simply put, investing in AI reduces cost per unit of output and creates economies of scale through reduced operating costs, lowering the cost of promotion and advertising (De Bruyn et al., 2020; Onyechi & Abeysinghe, 2009; Shaikh et al., 2021).

Automated processes: AI-powered bots can automate data entry and customer service, freeing up staff to work on more creative and effective marketing projects. AI can also improve business processes. This lets small businesses focus on growth and innovation, saving time and money (Alupo, Omeiza & Vernon, 2022; Haleem et al., 2022; Sithole & Ruhode, 2021).

Improved customer service: AI-powered chatbots can provide personalised, real-time support to customers 24/7. This can assist small businesses in developing stronger customer relationships, engagement and increasing customer satisfaction. Furthermore, AI-powered voice recognition technology can help customers find answers quickly and easily (Bettoni et al., 2021).

Increased profits: AI algorithms can recommend more likely-to-buy products and services. This can boost sales and profits by ensuring businesses are able to understand and meet consumer needs. AI-powered marketing tools can help small businesses target the right audiences and personalise campaigns for maximum ROI (Bettoni et al., 2021; Campbell et al., 2020; Dimitrieska, Stankovska & Efremova, 2018; Hansen & Bøgh, 2021).

Make smarter decisions: AI can find hidden patterns in analysing large datasets. This helps companies identify areas for improvement and better decision-making in product, pricing, distribution or promotion mix. AI can also predict events and customer needs. This can be especially useful for small businesses that can't afford market research (Bettoni et al., 2021; Campbell et al., 2020; Dimitrieska, Stankovska & Efremova, 2018; Hansen & Bøgh, 2021).

Increased productivity: Reduce time spent on repetitive tasks, allowing marketers to focus on other important tasks and implementation of campaigns (Schlögl et al., 2019).

Process optimisation and accuracy: Improve current work processes as well as improve the precision of marketing plans during the design, strategy and planning stages in business processes (De Bruyn et al., 2020; Huang & Rust, 2021; Regona et al., 2022).

Increased reliability: AI systems that exhibit satisfactory performance during their first stages but may thereafter demonstrate a propensity for errors (Passi & Vorvoreanu, 2022), although the technology can help produce results that are clear to all parties involved and enhance consistency and dependability (Campbell et al., 2020; Regona et al., 2022; Verma et al., 2021).

From the above, AI presents a wealth of opportunities for small businesses. It can help them to make better marketing decisions, automate processes, improve customer service and increase profits. As AI technology continues to evolve, small businesses should consider taking advantage of the many benefits it has to offer and adopt the technology within their business day-to-day marketing operations to leverage the opportunity available.

14.5.3 Challenges in AI Adoption for SMMEs

AI adoption also presents a range of challenges for small businesses. This section examines the challenges of adopting AI for small businesses.

Although AI is gaining popularity, the development in technology is also beginning to evaluate and support existing business models and procedures. About 70% of Chief Executive Officers (CEOs) in South Africa agree that investing in AI will increase productivity and competitiveness. Only 30% of these firms intend to make significant AI investments (Arakpogun et al., 2021).

Meanwhile, the most common challenges in AI adoption include: compatibility with current business processes and practices, highly specialised AI applications that require continuous algorithm training to detect data patterns, some industries may experience data acquisition issues, and AI platforms that require ongoing investment to ensure data accuracy and currency are the most common issues (Dwivedi et al., 2021; Regona et al., 2022). The security and dependability of vast amounts of data, the possibility of non-actionable tasks due to multi-point responsibility, and the difficulty of implementing AI in construction projects due to lack of standardisation are additional obstacles that will only slightly slow AI development in this industry (Dwivedi et al., 2021; Regona et al., 2022).

Based on literature review, the below highlights some challenges found common among SMMEs from a global perspective in adopting AI technologies.

Lack of AI adoption strategy/ applicable framework: The majority of businesses lack a clear strategy or the necessary skills to gather marketing data and incorporate AI into their regular marketing activities (Bettoni et al., 2021; Haleem et al., 2022).

Lack of resources to adopt AI: Implementing AI technologies requires expertise and resources. Small businesses may not have the funds or skills to hire AI experts or develop their own AI solutions. They may lack AI algorithm training and data. These may cause difficulty for small businesses to take advantage of the marketing opportunities AI provides (Bettoni et al., 2021; Onyechi & Abeysinghe, 2009; Shaikh et al., 2021). Lack of understanding of AI and the use of sufficient data collected to build algorithms, and the lack of understanding of AI tools and benefits can pose a challenge for adoption (Afolayan & de la Harpe, 2020; Dwivedi et al., 2021; Schlögl et al., 2019).

Concerns about reliability of technology: AI technologies are still evolving and may not produce the expected results. Furthermore, AI can be vulnerable to attacks, resulting in data breaches or service disruptions. This can result in customer distrust, lost sales and damage to a company's reputation (Campbell et al., 2020; De Bruyn et al., 2020; Huang & Rust, 2021).

Lack of data and security issues: Most often, businesses do not lack marketing data to feed into AI algorithms since continuous data to build patterns is critical. Data and security issues with the adoption of data protection laws, including the POPI Act in South Africa and the General Data Protection Regulation, also pose a challenge in data collection and compliance in Europe (Bettoni et al., 2021; Schlögl et al., 2019).

High investment costs of AI: To make AI solutions work, companies must invest in their infrastructure, however, these investments are perceived to be expensive for small businesses on a tight budget. There is also a lack of AI-related methods estimating the opportunity cost and return on investment of AI implementation (Bettoni et al., 2021).

Technology applicability and integration to current business: Companies struggle to find solutions that can be personalised to solve their problems as well as be easily integrated into their marketing activities or technologies with existing business systems (Afolayan & de la Harpe, 2020; Bettoni et al., 2021; Onyechi & Abeysinghe, 2009).

Ethical concerns: AI technologies have the potential to be used to monitor and make customer decisions, raising concerns about privacy and fairness. To ensure that their

AI systems are used responsibly, businesses must be aware of these issues and develop ethical AI policies (Dwivedi et al., 2021).

Technology complexity: AI tools and applications are still perceived to be too complex to use and due to technology having a lack of emotional intelligence and the ability to recognise and understand emotions in humans applicable to image recognition, voice analysis and text analysis which sometimes created a barrier for adoption (Bettoni et al., 2021; De Bruyn et al., 2020; Regona et al., 2022).

Lack of organisational and expert support: Corporate sponsorship to provide support in terms of required training, finance, among others to switch over technology. There is limited expert consultation to obtain information on new technology from technology experts in the industry (Shaikh et al., 2021).

New technology evaluation and risk management: There is currently a lack of an established framework for assessing new technologies and managing the associated risks during their implementation across diverse business operations (Afolayan & de la Harpe, 2020).

Return on investment and value of the new technology: The nature of the expected returns on the investment made with new technology has a potential impact on the adoption. This is also including the perceived value of investing in a new technology for marketing purposes (Afolayan & de la Harpe, 2020; De Bruyn et al., 2020).

Lack of skilled employees: Some businesses may need to invest in training and incorporating the investment in digital skills development and providing the required skills to ensure that they fully capitalise the value of adopting and using AI (Bettoni et al., 2021). Skills development in South Africa is a government-driven project that is strategically implemented through the Sector Education and Training Authorities (SETAs) (Viswanathan & Telukdarie, 2020).

Human resistance to technology: Some employees may not be ready for AI to take over or be integrated into their work. Employee attitude, buy-in and willingness to change from known routine practices and process changes to adapt and accept new technology (Afolayan & de la Harpe, 2020; Onyechi & Abeysinghe, 2009; Shaikh et al., 2021).

Biased artificial intelligence: Algorithm learning trends that may be unethical, prejudiced or create social biases which can be a hindrance in adopting new technology. This is seen in various examples where biases are identified with specific targeted consumer groups and AI systems displaying patterns of unfairness (De Bruyn et al., 2020; Regona et al., 2022).

Lack of government support: Various government-based programmes to help small business employees, managers and owners successfully adopt new technologies. (Dwivedi et al., 2021). In South Africa, enhancement of skills is prioritised through a specific programme initiated by the government and executed through the SETAs in a strategic manner (Matekenya & Moyo, 2022).

Fear of technology takeover: Fear of technology replacing the need for humans and thus leading to job losses and employee redundancy. As machines have improved, they can now perform more manual and cognitive activities, leading to public fears about job cuts and technological unemployment have grown (Arntz, Gregory & Zierahn, 2019).

From the above, small businesses face a range of challenges when considering the adoption of AI. However, there are solutions that can be considered to circumvent these challenges. Possible local relevant solutions are needed to address the challenges and ensure that solutions that are considered are effective.

14.6 Summary and Conclusion

The research provides a list of opportunities and challenges in adopting AI for marketing in SMMEs. The analysis in literature allowed to identify the existing challenges that pose a barrier in AI adoption and opportunities that may drive interest in businesses to adopt AI for marketing. In conclusion, small businesses face a range of challenges when adopting AI. Conversely, there are a number of potential solutions that can help to address these challenges. There is evidently a lot of potential that AI presents for small businesses to leverage. The research explored the various challenges and opportunities in adopting AI that could be relevant to South African SMMEs. The research is limited due to its being conceptual in nature as the selected methodology is that of a literature review. Empirical research is needed to validate the above AI adoption challenges and opportunities for small businesses. Future research should also examine AI adoption strategies, how to use AI in small businesses, how AI can create new business opportunities, and the ethical implications of AI adoption.

Bibliography

Afolayan, A. O., & de la Harpe, A. C. (2020). The role of evaluation in SMMEs' strategic decision-making on new technology adoption. *Technology Analysis & Strategic Management*, *32*(6), 697–710. https://doi.org/10.1080/09537325.2019.1702637.

AlSheibani, S., Cheung, Y., & Messom, C. (2018). Artificial intelligence adoption: AI-readiness at firm-level. https://researchmgt.monash.edu/ws/portalfiles/portal/273209396/254798983_oa.pdf (accessed April 2024).

Alupo, C. D., Omeiza, D., & Vernon, D. (2022). Realizing the potential of AI in Africa: It all turns on trust. In: M. I. A. Fereira, & M. O. Tokhi (Eds.), *Towards Trustworthy Artificial Intelligent Systems* (pp. 179–192). Springer. https://doi.org/10.1007/978-3-031-09823-9_13.

Arakpogun, E. O., Elsahn, Z., Olan, F., & Elsahn, F. (2021). Artificial intelligence in Africa: Challenges and opportunities. In: A. Hamdan, A. E. Hassanien, A. Razzaque, & B. Alareeni (Eds.), *The Fourth Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success* (pp. 375–388). Cham: Springer. https://doi.org/10.1007/978-3-030-62796-6_22.

Arntz, M., Gregory, T., & Zierahn, U. 2020. Digitalization and the Future of Work: Macroeconomic Consequences. *SSRN Electronic Journal*. 10.2139/ssrn.3411981.

Bettoni, A., Matteri, D., Montini, E., Gładysz, B., & Carpanzano, E. (2021). An AI adoption model for SMEs: A conceptual framework. *IFAC-PapersOnLine*, *54*(1), 702–708. https://doi.org/10.1016/j.ifacol.2021.08.082.

Bhorat, H., Asmal, Z., Lilenstein, K., & Van der Zee, K. (2018). SMMEs in South Africa: Understanding the constraints on growth and performance. Working paper 201802. https://commerce.uct.ac.za/sites/default/files/content_migration/commerce_uct_ac_za/1093/files/DPRU%2520WP201802.pdf (accessed April 2024).

Bushe, B. (2019). The causes and impact of business failure among small to micro and medium enterprises in South Africa. *Africa's Public Service Delivery and Performance Review*, *7*(1), 1–26. https://doi.org/10.4102/apsdpr.v7i1.210.

Campbell, C., Sands, S., Ferraro, C., Tsao, H. J., & Mavrommatis, A. (2020). From data to action: How marketers can leverage AI. *Business Horizons*, *63*(2), 227–243. https://doi.org/10.1016/j.bushor.2019.12.002.

Chen, H., Li, L., & Chen, Y. (2021). Explore success factors that impact artificial intelligence adoption on telecom industry in China. *Journal of Management Analytics*, *8*(1), 36–68. https://doi.org/10.1080/23270012.2020.1852895.

De Bruyn, A., Viswanathan, V., Beh, Y. S., Brock, J. K., & Von Wangenheim, F. (2020). Artificial intelligence and marketing: Pitfalls and opportunities. *Journal of Interactive Marketing*, *51*(1), 91–105. https://doi.org/10.1016/j.intmar.2020.04.0.

De Mauro, A., Sestino, A., & Bacconi, A. (2022). Machine learning and artificial intelligence use in marketing: A general taxonomy. *Italian Journal of Marketing*, *2022*(4), 439–457. https://doi.org/10.1007/s43039-022-00057-w.

Deloitte. (2022). 2022 connectivity and mobile trends. https://www2.deloitte.com/us/en/insights/industry/telecommunications/connectivity-mobile-trends-survey-2022.html#key-findings (accessed April 2024).

Dimitrieska, S., Stankovska, A., & Efremova, T. (2018). Artificial intelligence and marketing. *Entrepreneurship*, *6*(2), 298–304.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., & Eirug, A. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, Article 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002.

Haleem, A., Javaid, M., Qadri, M. A., Singh, R. P., & Suman, R. (2022). Artificial intelligence (AI) applications for marketing: A literature-based study. *International Journal of Intelligent Networks*, *3*, 119–132. https://doi.org/10.1016/j.ijin.2022.08.005.

Hansen, E. B., & Bøgh, S. (2021). Artificial intelligence and internet of things in small and medium-sized enterprises: A survey. *Journal of Manufacturing Systems*, *58*, Part B, 362–372. https://doi.org/10.1016/j.jmsy.2020.08.009.

Hollebeek LD, Sprott DE, Brady MK (2021) Rise of the machines? Customer engagement in automated service interactions. Journal of Service Research 24(1): 3–8

Huang, M., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, *49*(1), 30–50. https://doi.org/10.1007/s11747-020-00749-9.

Jarek, K., & Mazurek, G. (2019). Marketing and artificial intelligence. *Central European Business Review*, *8*(2), 46–55. https://doi.org/10.18267/j.cebr.213.

Kabalisa, R., & Altmann, J. (2021). AI technologies and motives for AI adoption by countries and firms: A systematic literature review. In: K. Tserpes, J. Altmann, J. A. Bañares, O. A. Ben-Yehuda, K. Djemame,

V. Stankovski, & B. Tuffin (Eds.), Lecture Notes in Computer Science: Vol. 13072. *Economics of Grids, Clouds, Systems, and Services* (pp. 39–51). Springer. https://doi.org/10.1007/978-3-030-92916-9_4.

Kofod-Petersen, A. (2012). How to do a structured literature review in computer science. Ver.0.1, 1. https://research.idi.ntnu.no/aimasters/files/SLR_HowTo2018.pdf (accessed April 2024).

Maroufkhani P., Iranmanesh M., Ghobakhloo M. (2023). Determinants of big data analytics adoption in small and medium-sized enterprises (SMEs). *Industrial Management & Data Systems*, 123, 278–301.

Matekenya, W., & Moyo, C. (2022). Innovation as a driver of SMME performance in South Africa: A quantile regression approach. *African Journal of Economic and Management Studies*, *13*(3), 452–467. https://doi.org/10.1108/AJEMS-06-2021-0306.

Matta, V., Bansal, G., Akakpo, F., Christian, S., Jain, S., Poggemann, D., Rousseau, J., & Ward, E. (2022). Diverse perspectives on bias in AI. *Journal of Information Technology Case and Application Research*, *24*(2), 135–143. https://doi.org/10.1080/15228053.2022.2095776.

Morgan, N. A. (2012). Marketing and business performance. *Journal of the Academy of Marketing Science*, *40*(1), 102–119. https://doi.org/10.1007/s11747-011-0279-9.

Onyechi, G. C., & Abeysinghe, G. (2009). Adoption of web based collaboration tools in the enterprise: Challenges and opportunities. Paper presented at the 2009 International Conference on the Current Trends in Information Technology (CTIT), 1–6.

Passi, S & Vorvoreanu, M. 2022. Overreliance on AI: Literature Review. Microsoft Technical Report MSR-TR-*2022*-12

Radhakrishnan, J., & Chattopadhyay, M. (2020). Determinants and barriers of artificial intelligence adoption – A literature review. Paper presented at the International Working Conference on Transfer and Diffusion of IT, 89–99.

Regona, M., Yigitcanlar, T., Xia, B., & Li, R. Y. M. (2022). Opportunities and adoption challenges of AI in the construction industry: A PRISMA review. *Journal of Open Innovation: Technology, Market, and Complexity*, *8*(1), Article 45. https://doi.org/10.3390/joitmc8010045.

Schlögl, S., Postulka, C., Bernsteiner, R., & Ploder, C. (2019). Artificial intelligence tool penetration in business: Adoption, challenges and fears. Paper presented at the International Conference on Knowledge Management in Organizations, 259–270.

SEDA. (2022). SMME quarterly update Q3 2022. https://www.seda.org.za/Publications/Publications/SMME%20Quarterly%202022-Q3%20(005).pdf (accessed April 2024).

Shaikh, D., Ara, A., Kumar, M., Syed, D., Ali, A., & Shaikh, M. Z. (2021). A two-decade literature review on challenges faced by SMEs in technology adoption. *Academy of Marketing Studies Journal*, *25*(3). https://www.abacademies.org/articles/A-two-decade-literature-review-on-challenges-faced-by-1528-2678-25-3-398.pdf (accessed April 2024).

Sithole, S. S., & Ruhode, E. (2021). Cloud computing adoption: Opportunities and challenges for small, medium and micro enterprises in South Africa. https://digitalknowledge.cput.ac.za/bitstream/11189/8311/1/Cloud%20Computing%20Adoption.pdf (accessed April 2024).

Statistics South Africa (Ed). (2022). Quarterly labour force survey; quarter 4: 2021 No. 11). https://www.statssa.gov.za/publications/P0211/P02114thQuarter2022.pdf (accessed April 2024).

Taljaard, T., & Gerber, A. (2022). The preparation of South African companies for the impact of artificial intelligence. Paper presented at the Southern African Conference for Artificial Intelligence Research, 348–367.

Telukdarie, A., Philbin, S., Mwanza, B. G., & Munsamy, M. (2022). Digital platforms for SMME enablement. *Procedia Computer Science*, 200, 811–819. https://doi.org/10.1016/j.procs.2022.01.278.

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.

Venkatesh, V. (2022). Adoption and use of AI tools: A research agenda grounded in UTAUT. *Annals of Operations Research*, 308(1–2), 641–652. https://doi.org/10.1007/s10479-020-03918-9.

Verma, S., Sharma, R., Deb, S., & Maitra, D. (2021). Artificial intelligence in marketing: Systematic review and future research direction. *International Journal of Information Management Data Insights*, *1*(1), Article 100002. https://doi.org/10.1016/j.jjimei.2020.100002.

Viswanathan, R & Telukdarie, A. 2020. A systems dynamics approach to SME digitalization IEEE International Symposium on Multimedia.

We Are Social. (2022). Global overview report. https://datareportal.com/reports/digital-2022-global-overview-report.

Part 7: **Forecasting**

Katharina I. Köstner, Bàrbara Llacay, David Alaminos

Chapter 15
Forecasting Brent Oil Volatility: DeepAR vs LSTM

Abstract: The recent international crises have revealed the economy's vulnerability to fluctuations in commodity prices. Expectations on the volatility of the current price mark strategic decisions of companies and investors, especially in the case of oil. Given its importance, modelling with computational techniques has intensified in recent years. This study models the prediction of Brent oil volatility with Deep Autoregressive Recurrent Networks (DeepAR), an algorithm drawing from traditional autoregressive models. Our results demonstrate an improvement in the classification results and error levels with the DeepAR after comparing to Long Short-Term Memory (LSTM) and previous similar works, with a level range of 0.0007–0.0057 in root mean squared error (RMSE) terms. The application of the methodology offers a new alternative in volatility prediction to different interest groups, such as companies, institutional and individual investors and public institutions, as well as the possibility of broadening its applicability to the volatility of other assets.

Keywords: Deep Learning, Volatility, Oil Price, LSTM, DeepAR

15.1 Introduction

Brent oil is currently among the most volatile commodities. The impact of the COVID-19 pandemic, followed by the deterioration of geopolitical stability in the aftermath of the Russian invasion of Ukraine, has sent markets spinning. To illustrate, after a historic low of around $18.38 per barrel in March 2020, Brent oil reached a 10-year high of $122.71 per barrel in 2022. Considering the uncertain economic outlook for 2023, it is reasonable to assume that these price swings will continue. This is an issue for multiple reasons. Oil remains one of the most relevant commodities in global markets (Bourghelle, Jawadi & Rozin, 2021; Chatziantoniou et al., 2021). The Granger hypothesis of economic contraction as a result of oil price hikes has been called into question, yet

Katharina I. Köstner, PhD Programme in Business, Universitat de Barcelona, Barcelona, Spain, e-mail: 31koestk7@alumnes.ub.edu
Bàrbara Llacay, Department of Business, Universitat de Barcelona, Barcelona, Spain, e-mail: bllacay@ub.edu
David Alaminos, Department of Business, Universitat de Barcelona, Barcelona, Spain, e-mail: alaminos@ub.edu

https://doi.org/10.1515/9783110790320-015

there is undoubtedly a correlation between oil price developments and economic performance. As an energy source, as well as an input product in various stages of the manufacturing process, oil is not fully replaceable yet and as such has strong implications for economic outcomes. For this reason, oil price volatility has significant spillover effects on other markets. This is exacerbated by the fact that oil price volatility is increased by the financialisaton of oil markets – especially the introduction of oil exchange-traded funds (ETFs) – which has led to increased speculation with the commodity. Simultaneously, accurate volatility forecasts are essential as policymakers, investors and hedgers base decisions on them. Improving forecasting techniques is therefore a fruitful endeavour (Carpio, 2019).

However, the forecasting of volatility is not trivial. While there are several econometric models that approach this issue, these models suffer from estimation constraints as well as problems with generalisation error (Sehgal & Pandey, 2015; Zhang, Ma & Wei, 2019). The improved capabilities of self-learning algorithms provide a new angle (Chiroma et al., 2016; Lu et al., 2022). The popularity of deep learning models is easily explained. The continuing automation of financial processes constitutes a competitive advantage, potentially enhancing model accuracy and improving compliance processes. Deep learning algorithms are successfully employed in credit risk estimation and fraud detection. In the context of financial modelling, the adoption of deep learning techniques has increased significantly in past years. Deep learning algorithms constitute universal function approximators, that is, a deep learning model with an optimal number of neurons can approximate any functional form. This is a considerable advantage over the, at times, rigid assumptions of conventional econometric models. Moreover, in the context of forecasting tasks, an advantage of deep learning models is that the underlying variable distribution need not be known, but can be learned by the model, which decreases the potential for erroneous models based on inaccurate assumptions.

However, the application of deep learning also presents new challenges. How can deep learning techniques be applied to volatility forecasting? Which algorithms are best suited for the task? What are their limitations? How do they perform in different contexts? To address this, this chapter utilises a recent deep learning algorithm specifically developed for time series forecasting that, to the best of our knowledge, has not been applied to the forecasting of volatility: Deep Autoregressive Recurrent Networks (DeepAR) (Salinas et al., 2020). The DeepAR algorithm has received significant attention in deep learning research. It delivers probabilistic forecasts for multiple series and interestingly is based on the autoregressive integrated moving average (ARIMA) model for time series analysis (Salinas et al., 2020). The algorithm has been shown to handle time series forecasting well and could potentially outperform traditional RNNs (Zhang, Wang & Wang, 2020).

As such, this chapter applies DeepAR to the forecasting of Brent price volatilities, with a focus on periods of high market uncertainty. This chapter contributes to the literature by assessing if utilising deep learning is a valuable technique to forecast extreme values and assess if novel deep learning algorithms can yield forecasting improvement in excess of traditional RNNs.

The rest of the chapter is composed as follows. Sections 16.2 and 16.3 provide a review of volatility forecasting methods and oil price volatility. In Section 16.4, the methodology is described. Section 16.5 details the data and sample involved in the research. Section 16.6 discusses model training and hyperparameters. Section 16.7 points out the results and findings obtained. Section 16.8 concludes the chapter.

15.2 Volatility Forecasting Techniques

Forecasting volatility is not entirely straightforward and therefore makes for an interesting research object. The first issue presenting itself is that volatility in and of itself cannot be measured directly. Generally, it is defined as the dispersion of values around the mean, but it cannot be directly observed. Engle and Patton (2001) have studied stylised facts about volatility. Firstly, they find that volatility tends to cluster. As such, periods of low volatility are likely followed by periods of low volatility and vice versa. Secondly, volatility is mean reverting, that is, the volatility of a certain stock or commodity tends to revert back to long-term trends after short-term deviations. A potential explanation is that long-term volatility reflects the underlying characteristics of the asset or commodity in question. Moreover, volatility appears to be asymmetrically impacted by the arrival of new information.

The main target, therefore, is to find a volatility model to capture these effects. To this end, several models have been developed in the literature. Poon and Granger (2003) provide a detailed review. One class of models constitutes stochastic volatility models, originating in option pricing and first proposed by Hull and White (1987). This contribution was extended by Heston (1993). However, as these models assume continuous values, discretisation error becomes a problem with real-world infrequent data, and they are rarely used in forecasting as a result.

A more practicable class of models focuses on autoregressive (AR) models. Traditional AR models do not account for the heteroscedastic term structure of volatility. This was first considered by Engle (1982) with the development of the Autoregressive Conditional Heteroskedasticity (ARCH) model. Subsequently, Bollerslev (1986) extended this model to develop the generalised autoregressive heteroskedasticity (GARCH) model, which is by far the most widely known and used volatility forecasting model to date. Both ARCH and GARCH are capable of capturing heteroskedasticity, clustering and mean reversion, however, they cannot capture an asymmetric distribution of returns. Therefore, several extensions to the GARCH model have been introduced in the literature, such as Exponential GARCH (Nelson, 1991), Quadratic GARCH (Sentana, 1995), Threshold GARCH (Zakoian, 1994) and FIGARCH (Fractionally integrated general autoregressive conditional heteroskedasticity model) (Baillie, Boi|erslev & Mikkelsen, 1996).

Another notable addition to the volatility forecasting literature was made after the observation of the distribution of realised volatility, made feasible by the introduction of high-frequency trading which enabled better data collection techniques (Andersen et al., 2001). The adoption of vector autoregressive stochastic volatility was a result, as was the introduction of Corsi's heterogeneous autoregressive (HAR) model (Corsi, 2009).

While these models have been proven to work well in practice and are widely adopted by the financial services industry, they do exhibit several drawbacks. One is that forecasting tasks need to be calibrated on the dataset with maximum likelihood estimation. Consequently, they generally perform well in-sample but poorly out-of-sample. Besides generalisation issues, they also assume a certain functional form. Since deep learning algorithms do not suffer from this limitation and have been developed to minimise generalisation error, it is reasonable to assume that their adoption can yield forecasting improvements. To summarise, a good volatility model should (a) capture the stylised facts of volatility well, (b) deal with extreme values, (c) deal with sparse, non-continuous data and (d) generalise well to unseen data.

The literature on deep learning algorithms for volatility forecasting is currently surprisingly sparse albeit a growing field. Ge et al. (2022) provide a detailed review. There are three basic types of deep learning algorithms: Multilayer Perceptrons (MLPs), Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). CNNs were created for image recognition and contain convolutional layers that are designed to classify pixels into certain areas to facilitate learning. By the nature of convolutional layers, these are not well suited to predict time series data for volatility. As a result, there is only one paper utilising a hybrid CNN-RNN for forecasting volatility (Vidal and Kristjanpoller, 2020). By contrast, MLPs have been used frequently in the literature, particularly hybridised versions as in Kristjanpoller and Hernández (2017), Ramos-Pérez, Alonso-González and Núñez-Velázquez, (2019) and Ribeiro et al. (2021). However, the MLP assumes data to be identically and independently distributed and as such are not utilising the time series structure of volatility data. By contrast, RNNs were specifically designed for this task and have been applied to the field of forecasting oil price volatility. The following section gives an overview of the current state of the art regarding oil price volatility forecasts. The details of the Long Short-Term Memory (LSTM) and DeepAR algorithms are presented in Section 16.4.

15.3 Oil Price Volatility Forecasting

To date, in line with the general literature on volatility forecasting, most researchers have focused on GARCH models when assessing oil volatility. Lang and Auer (2020) provide a comprehensive review. In recent years, machine learning methods have received increasing attention in the field. For example, Niu and Zhao (2021) present a hybrid forecasting model based on variational mode decomposition (VMD) and kernel extreme learning machine (KELM) to forecast the daily prices and 7-day volatility of crude oil. Gupta and Pierdzioch (2022) use machine learning algorithms to compare the contribution of aggregated versus disaggregated metrics of policy, stock market and geopolitical events uncertainties when forecasting the future realised volatility of oil price returns. Tissaoui et al. (2022) compare the performance of two machine-learning algorithms to forecast crude oil volatility before and during the COVID-19 pandemic. Zhang, Wahab and Wang (2023) use a range of machine learning techniques to predict aggregate oil market volatility by using a large macroeconomic database.

However, there are still few contributions that deal with the application of deep learning techniques to oil price volatility. Kristjanpoller and Minutolo (2016) use a hybrid artificial neural network in combination with a GARCH model (ANN-GARCH) to forecast oil return volatility. Lu et al. (2022) use both traditional machine learning models and neural networks to model oil price volatility building on a number of macroeconomic and financial predictors. Kamdem, Essomba and Berinyuy (2020) use an LSTM model to analyse the impact of COVID-19 on the forecasting of the volatility of different commodities, including crude oil. Verma (2021) builds a hybrid GARCH-RNN model to forecast the volatility of oil futures. Jiao et al. (2021) utilise an LSTM model and text-based sentiment analysis to forecast crude oil volatility. Li et al. (2021) also focus on the role of text-based information in oil volatility forecasting. Rakpho, Yamaka and Phadkantha, (2022) compare different hybrid ANN-GARCH models to forecast the volatility of crude oil, ethanol and natural gas. However, few of these papers include periods with extreme volatility events or examine which model is best suited to the forecasting task in extreme market environments as has been observed this year. The only exception is the paper by Cheng et al. (2022), where the authors propose an integrated framework consisting of complementary ensemble empirical mode decomposition (CEEMD), ARIMA and support vector machine (SVM) to forecast the volatility of crude oil by considering the impact of extreme events. In fact, the decomposition provided by the CEEMD component shows that extreme events have the most important effect in determining the prices of crude oil.

Many of the algorithms applied are somewhat outdated. For this reason, we use DeepAR (Salinas, Flunkert & Gasthaus, 2017) to forecast oil price volatility. Within the field of financial modelling, DeepAR has only been used to forecast stock indices in combination with algorithms to extract information from daily news (Barbaglia, Consoli & Wang, 2021; Consoli et al., 2022). Below, we give an introduction to the functioning of RNNs and the models applied.

15.4 Models Employed

An RNN is a deep learning network specifically designed to deal with time series data. While other deep learning networks, such as the MLP, show good accuracy and robustness in forecasting, they assume data to be independent and identically distributed and do not consider the time-ordered structure of data required for forecasting. RNNs have been developed specifically for this task. This is achieved by incorporating time-lagged hidden parameters. Another key feature of RNNs is the instance of so-called encoders and decoders, which results in their ability to transform a vector of samples into a vector of predictions of a different size.

One issue arising from this type of architecture is the increase in the number of input parameters, which makes RNNs susceptible to the vanishing or exploding gradient problem. Because of the increased number of input variables, the amount of required matrix multiplications increases, which in turn increases the probability of parameters approaching zero or increasing indefinitely, resulting in training breakdown. This was first solved by Hochreiter and Schmidhuber (1997) with the development of the LSTM algorithm. The added features that prevent the gradient problem are firstly the introduction of element-wise multiplication, which avoids the issue of full matrix multiplication and exponential increase and decrease of training parameters. Secondly, an update to the algorithm introduced so-called forget gates that enable the algorithm to drop batches of data that are irrelevant to training the algorithm when no further improvement is made.

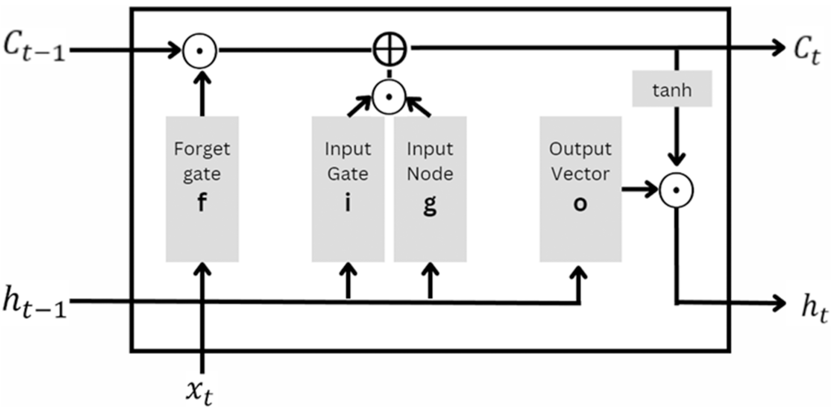


Figure 15.1: LSTM memory cell. (Compiled by authors.)

Figure 15.1 shows a scheme of an LSTM cell, where the symbols $\odot$ and $\oplus$ denote element-wise multiplication and summation, respectively. The forget gate f_t is determined by the previous hidden state h_{t-1} and the vector of input samples x_t. Element-wise multiplication with the previous cell state C_{t-1} determines the degree to which the previous cell state is considered in the current cell state. The previous hidden state h_{t-1} and the input variable x_t also determine the input gate i_t and input node g_t, as well as the output vector o_t.

Following the notation of Hochreiter and Schmidhuber (1997), cell state C_t is then defined as follows:

$$C_t = (C_{t-1} \odot_t) \oplus (i_t \odot g_t) \tag{1}$$

where:

$$f_t = \sigma\left(W_{xf} x_t + W_{hf} h_{t-1} + b_f\right) \tag{2}$$

$$i_t = \sigma(W_{xi} x_t + W_{hi} h_{t-1} + b_i) \tag{3}$$

$$g_t = tanh\left(W_{xg} x_t + W_{hg} h_{t-1} + b_g\right) \tag{4}$$

That is, the cell state is the element-wise sum of the element-wise product of the previous cell state C_{t-1} and the forget gate f_t and the element-wise product of an input gate i_t and the input node g_t. Based on the cell state, the hidden units are defined by:

$$h_t = o_t \odot \tanh(C_t) \tag{5}$$

The output vector is then calculated as follows:

$$o_t = \sigma(W_{xo} x_t + W_{ho} h_{t-1} + b_o) \tag{6}$$

To date, the LSTM algorithm is one of the most popular models in the forecasting literature, particularly in the field of economic variable forecasting, so this model is used as a baseline to compare to the DeepAR model.

The defining feature of DeepAR is its ability to generate probabilistic forecasts. To this end, the algorithm calculates a conditional probability density function from which samples of the target variable $z_{i,t}$ are then drawn. Figure 15.2 shows the simplified architecture of the model following the notation of Salinas et al. (2020).

The input variables consist of a vector of previous values of the target value $z_{i,t-1}$, a vector of control variables $x_{i,t}$ and a vector of the previous hidden states $h_{i,t}$. As the indexing of the variables indicates, DeepAR allows simultaneous forecasting of multiple time series. It is also able to consider control variables, for which future outputs are known, such as for example, days of the week or public holidays.

The goal here is to utilise information from other datasets to extrapolate data for the series forecast in question. The variables are then passed to three stacked LSTM networks, from which the current hidden state $h_{i,t}$ is determined:

$$h_{i,t} = h(h_{i,t-1}, z_{i,t-1}, x_t, \Theta) \tag{7}$$

The probability density function is then calculated by

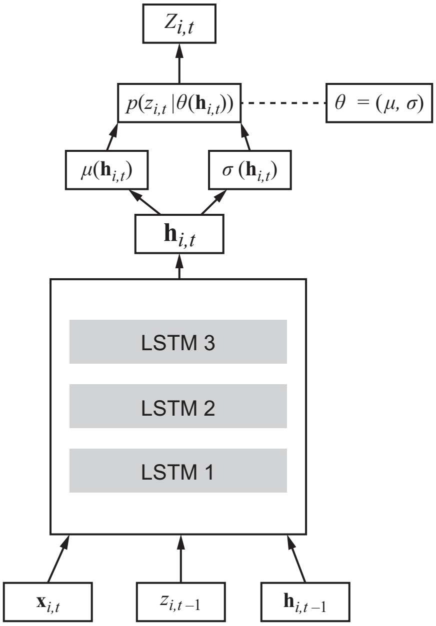


Figure 15.2: DeepAR architecture. (Compiled by authors.)

$$p\_G\left(z|\mu,\sigma\right) = \frac{\left(2\pi\sigma^2\right)^{-\frac{1}{2}}\exp((-z-\mu)^2}{2\sigma^2}\Bigg) \tag{8}$$

and parameterised by mean $\mu(h_{i,t})$ and standard deviation $\sigma(h_{i,t})$:

$$\mu(h_{i,t}) = w_\mu^T h_{i,t} + b_\mu \tag{9}$$

$$\sigma(h_{i,t}) = \log(1 + \exp(w_\sigma^T h_{i,t} + b_\sigma)) \tag{10}$$

Salinas et al. (2020) note that the probability density function should reflect statistical properties of the data. Since normal distribution of financial returns is a common assumption, it seems reasonable that a Gaussian likelihood function should work with our dataset. However, an examination of the effect of different density functions on model accuracy and generalisation could be a worthwhile addition in future research.

15.5 Data Description

This analysis is carried out on Brent crude oil prices in USD. The data has been collected from the U.S. Energy Information Administration, which makes its datasets freely available via their API.

The timeframe examined is from 20 May 1987 to 1 January 2023.

Because data is available in daily intervals only and values are therefore discrete, the volatility measure applied is historical volatility defined as

$$HV = \sqrt{\frac{1}{N} * \sum_{t=\tau_1}^{\tau_2} (r_t - \bar{r}_{\tau_1,\tau_2})^2}, \tag{11}$$

where $r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$ is the log return of Brent prices and $\bar{r}_{\tau_1,\tau_2} = \frac{1}{N} * \sum_{t=\tau_1}^{\tau_2} r_t$ is the mean of log returns. Historical volatility is calculated for a 21-day rolling window, to capture volatility effects for the past month, 21 being the average number of working days per month.

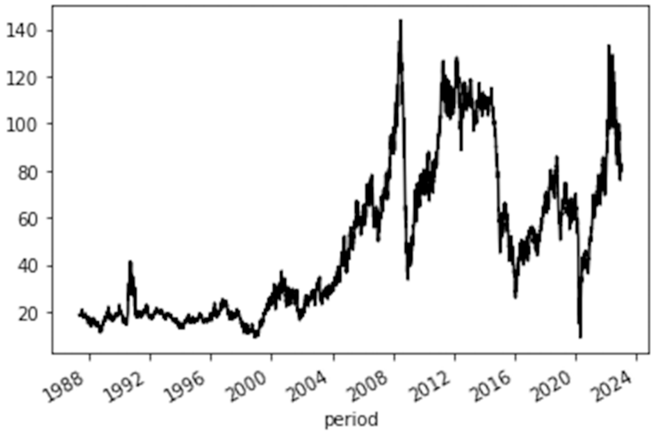


Figure 15.3: Brent Price in USD. (Compiled by authors.)

Figure 15.3 shows the Brent price development over the entire dataset from 1987 to present day. The price spikes preceding the financial crisis of 2007/8, following the period of economic recovery in the 2010s and finally the Russian invasion of Ukraine are clearly visible, as are the price drops following the financial crisis as well as the COVID-19 pandemic.

Taking log returns into account (Figure 15.4), the extreme oil price developments of the last three years become more apparent.

Figure 15.5 shows the volatility of the Brent price for the selected time period. Engle and Patton's (2001) findings of mean-reverting volatility are confirmed. Interestingly, the increase in volatility during the beginning of the COVID-19 pandemic and the succeeding economic slowdown are more pronounced than the increase in volatility following the war in Ukraine. This is consistent with the well-documented observation that volatility tends to increase more sharply in bear markets.

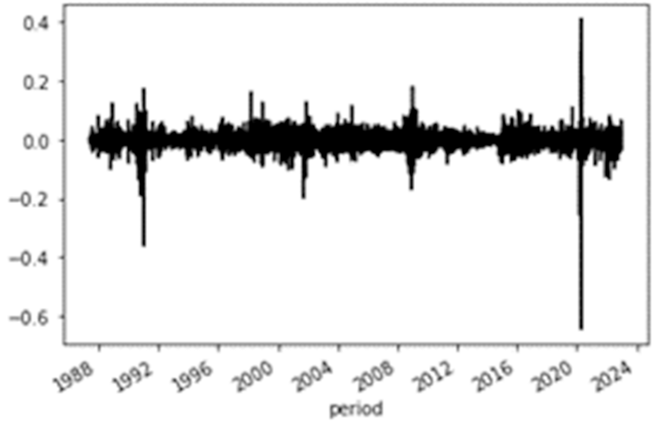


Figure 15.4: Brent log returns. (Compiled by authors.)

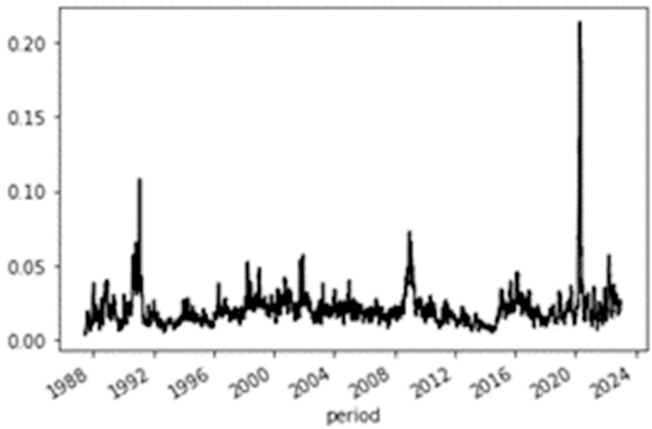


Figure 15.5: Brent Volatility. (Compiled by authors.)

To assess the ability of our models to forecast these spikes, it is necessary to define periods of high volatility. To this end, periods of high volatility are defined as periods, where volatility is in excess of one and a half standard deviations above the mean of the series. By this definition, eight periods of increased Brent price volatility can be identified as highlighted in Figure 15.6.

Observed volatility is markedly increased in early 2020 as well as in early 2022. Additionally, the dataset encompasses a sufficient number of volatility shocks to assume a well-trained algorithm can learn patterns of jumps from our training dataset. Having identified our forecasting target, we turn to the discussion of model selection.

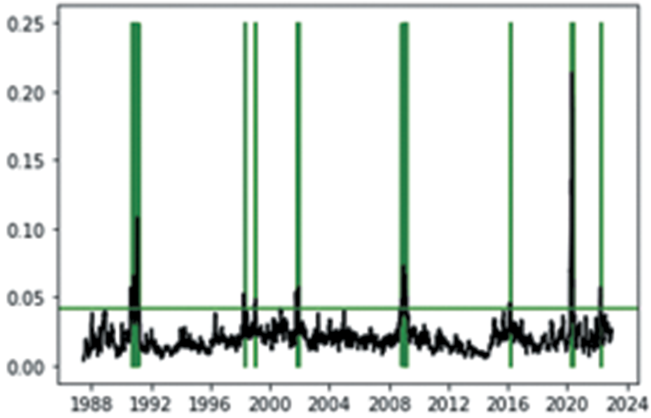


Figure 15.6: Periods of high volatility. (Compiled by authors.)

15.6 Model Training and Hyperparameters

The baseline model selected for our chapter is a vanilla LSTM with one layer and eight neurons. The model includes two dense layers with linear activation function. The model is trained with Adam optimiser with mean squared error (MSE) loss function, defined below. One advantage of the Adam optimiser is its ability to update the learning rate over time, thus decreasing model dependency on this hyperparameter. The initial learning rate is 0.01. The algorithm is trained over 50 epochs with a batch size of 50 data points. During training, 5% of samples are set aside for validation of the model. To find optimal hyperparameters, a grid search with various input parameters has been performed. The selected parameters are summarised in Table 15.1. While the LSTM is trained on a rolling window of 21-step, forecasts are made as one-step ahead forecasts. In order to achieve predictions of 7, 14 and 21 days, we forecast volatility one step ahead and the obtained parameter is then added to the test set and a new prediction is made based on the predicted value. This process is continued for the respective forecasting window so that we arrive at a 7-, 14- and 21-day prediction window as shown in Figure 15.7.

Our main model is based on the DeepAR architecture proposed by Salinas et al. (2020). It is comprised of three stacked LSTM layers with 30 nodes. The learning rate is set to 0.001. During initial testing, we experimented with different learning rates. We find that for the DeepAR model, the learning rate combined with the number of LSTM nodes is the most important hyperparameter. While higher learning rates tend to underfit the data and produce forecasts that largely mirror the series average, very small learning rates overfit the model, which decreases the model's ability to forecast abnormal volatility increases.

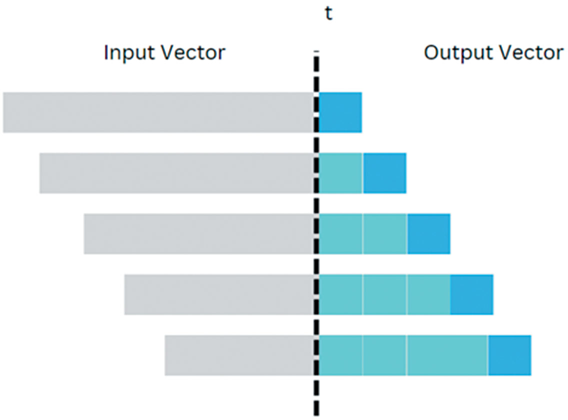


Figure 15.7: LSTM rolling window. (Compiled by authors.)

Table 15.1: Hyperparameters. (Compiled by authors.)

	DeepAR	LSTM
Batch size	50	50
Number of epochs	30	50
Learning rate	0.001	0.01 (initial)
# of LSTM layers	3	1
# of LSTM nodes	30	8
Activation function	Normal distribution loss	Mean squared error

The model is trained for 30 epochs with a batch size of 50 data points. The loss function selected is Pytorch's default normal distribution loss, to account for the probabilistic nature of the DeepAR model. Early stopping is applied, enabling the algorithm to stop training when model accuracy stagnates. The selected hyperparameters are largely based on the original model by Salinas et al. (2020), such as the number of layers and activation function. The learning rate and number of LSTM nodes were found in a series of experiments. The optimal model parameters are shown in Table 15.1.

The DeepAR algorithm is designed to make multiple day-ahead forecasts, so there is no need for iteration and the algorithm automatically uses the obtained one-step ahead forecasts for its predictions (Figure 15.8).

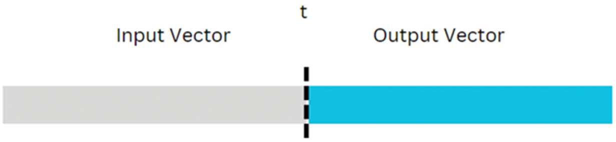


Figure 15.8: DeepAR forecasting window. (Compiled by authors.)

Model performance is evaluated based on MSE, root mean squared error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) as defined below:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_{actual} - y_{predict}\right| \tag{12}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_{actual} - y_{predict}}{y_{actual}}\right| \tag{13}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_{actual} - y_{predict}\right)^2 \tag{14}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{actual} - y_{predict}\right)^2} \tag{15}$$

where y_{actual} are the observed values and $y_{predict}$ are the values predicted by our model.

Training for both models is performed on data up to 2020. The model is then evaluated on a test set comprised of samples from the beginning of 2020 to the beginning of 2023, of which a window of 42 days is selected for evaluation. For the forecasts of volatility spikes in 2020 and 2022, respectively, the first day of volatility crossing the pre-defined threshold of 1.5 standard deviations above the mean is selected. The preceding window is then fed to the algorithm, which then forecasts the volatility spike. Results are presented in Sections 16.7.1. and 16.7.2, respectively.

15.7 Overall Model Performance Across Rolling Windows

We begin by evaluating general model performance by feeding the model input data of the past 21 days. A forecast is then made for the following 21-, 14- and 7-day window. This is done to test model and forecast robustness more thoroughly. The results in terms of the above error metrics are presented in Table 15.2.

DeepAR clearly outperforms LSTM over all error metrics. Unsurprisingly, 14- and 7-day ahead forecasts are more accurate than the 21-step ahead forecast. It should be noted that the 7-day ahead forecast appears to be less accurate than the 14-day ahead forecast, which could be explained by the window selected for evaluation. An example output of our forecast with DeepAR is shown in Figures 15.9–15.12. It should be noted that while calibrating the LSTM, we experimented with different model specifications. If one-step ahead forecasts are performed, the algorithm produces very accurate forecasts with significantly lower errors. However, one may question the usefulness of one-step ahead forecasting in real-world applications, especially as the LSTM still requires at least one time step to adjust to volatility shocks in our experiments.

Table 15.2: Error metrics for volatility forecasts with 21 sample inputs. (Compiled by authors.)

	MAE	MAPE	MSE	RMSE	MAE	MAPE	MSE	RMSE	MAE	MAPE	MSE	RMSE
	21 days				14 days				7 days			
DeepAR	0.0026	0.1085	0.0723	0.0027	0.0006	0.0222	0.0055	0.0007	0.0009	0.0361	0.0155	0.0012
LSTM	2.0122	87.1906	0.0135	0.1159	0.9144	38.1237	0.0035	0.0593	0.8295	36.1271	0.0033	0.0575

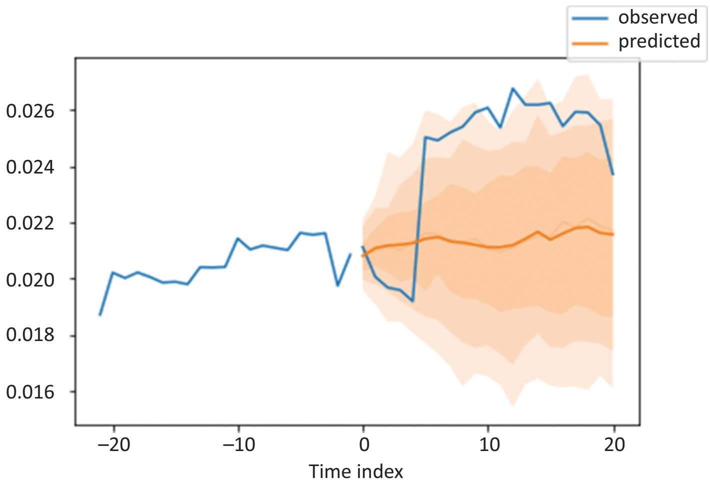


Figure 15.9: 21-day ahead forecast with 21-step input – DeepAR. (Compiled by authors.)

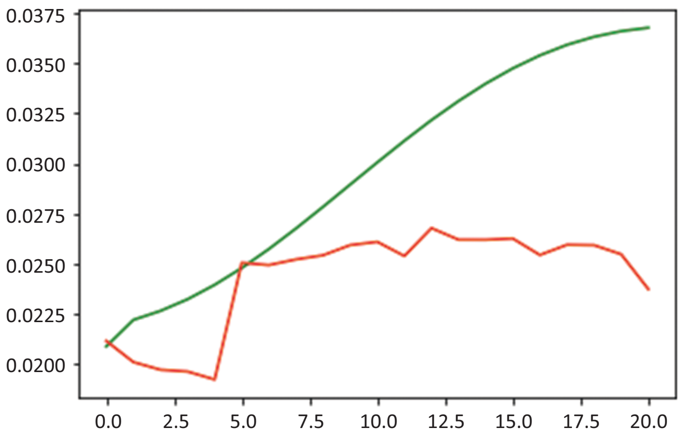


Figure 15.10: 21-step ahead forecast with 21-step input – LSTM, observed values in red, forecast in green. (Compiled by authors.)

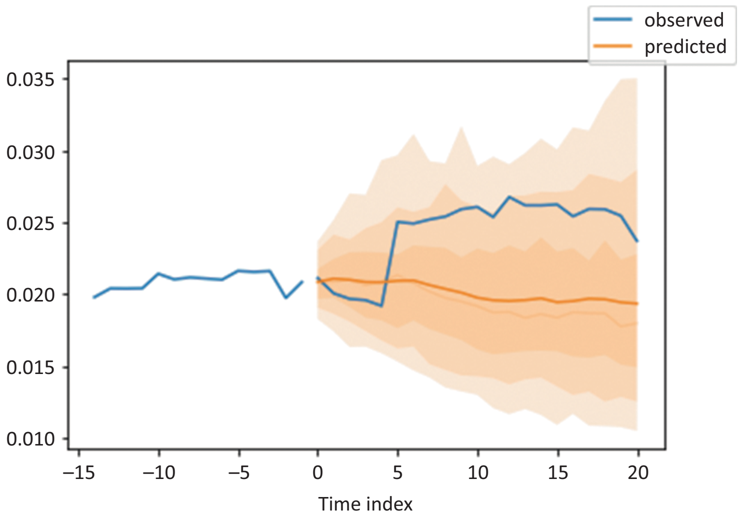


Figure 15.11: 14-day ahead forecast with 21-step input – DeepAR. (Compiled by authors.)

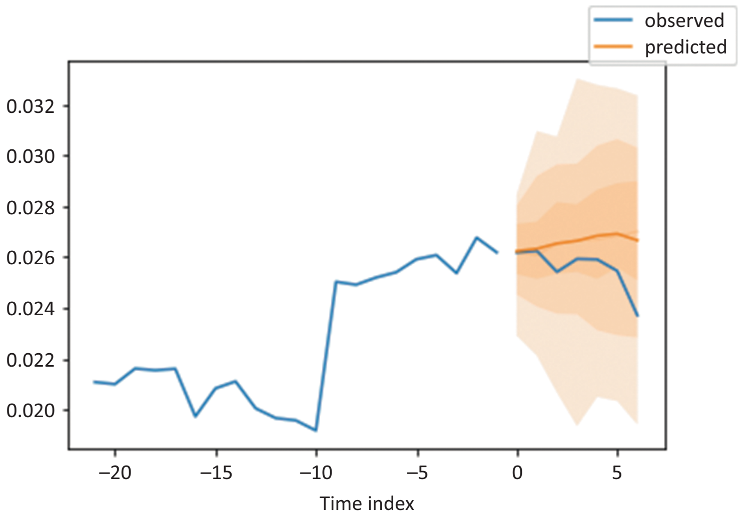


Figure 15.12: 7-day ahead forecast with 21-step input – DeepAR. (Compiled by authors.)

In the second step, the models are compared based on an input window of 14 steps. Again, DeepAR outperforms LSTM. In line with our expectations, there is a slight increase in the size of the error terms because of the restriction of input data. Again, error metrics decrease the fewer days ahead are forecast.

Table 15.3: Error metrics for volatility forecasts with 21 sample input. (Compiled by authors.)

	MAE	MAPE	MAE	MAPE	MAE	MAPE
	21 days		14 days		7 days	
DeepAR	0.0050	0.1964	0.0007	0.0277	0.0005	0.0192
LSTM	1.1746	48.4837	1.3368	52.6317	1.3156	53.0924
	MSE	**RMSE**	**MSE**	**RMSE**	**MSE**	**RMSE**
	21 days		14 days		7 days	
DeepAR	0.3209	0.0057	0.0066	0.0008	0.0052	0.0007
LSTM	0.3478	0.5897	0.0076	0.0873	0.0132	0.1149

15.7.1 Forecasting the Volatility Increases of 2020 and 2022

In the last step, we assess the forecasting accuracy of our models on volatility shocks. The relevant time periods are selected based on our definition above. Table 15.4 shows our selected error metrics. Figures 15.13 and 15.14 plot our forecasts against observed volatilities. Our model performs better while forecasting the volatility spike of February 2022. This is likely explained by the fact that volatility increased at a lower rate than it did in March 2020. This is in line with the observation of volatility spikes being more pronounced in market downturns and this can possibly be explained by the way investor beliefs about economic conditions were formed. The LSTM model we trained was not able to render accurate forecasts of the volatility spikes.

Table 15.4: Error metrics for forecasts during volatility events. (Compiled by authors.)

	MAE	MAPE	MSE	RMSE	MAE	MAPE	MSE	RMSE
	Mar-20				Feb-22			
DeepAR	0.0196	0.2290	0.0005	0.0220	0.0151	0.2888	0.0002	0.0154
LSTM	62.8775	367.2119	9.5822	3.0955	8.7165	229.2072	0.1796	0.4238

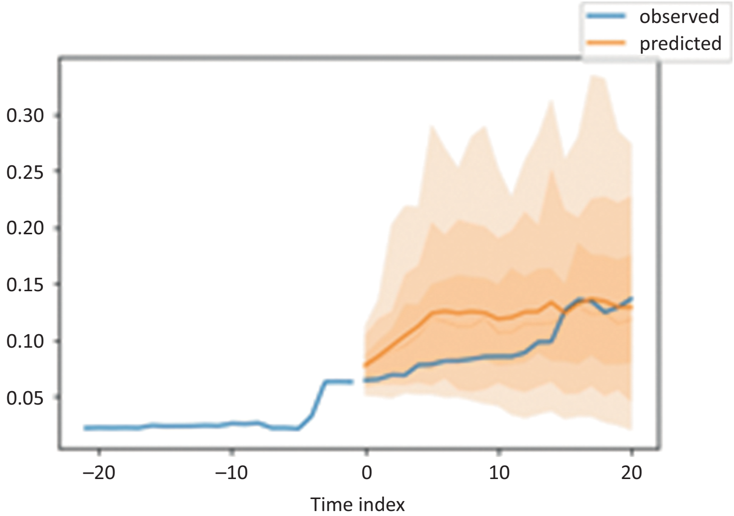


Figure 15.13: Volatility forecasts March 2020 – DeepAR. (Compiled by authors.)

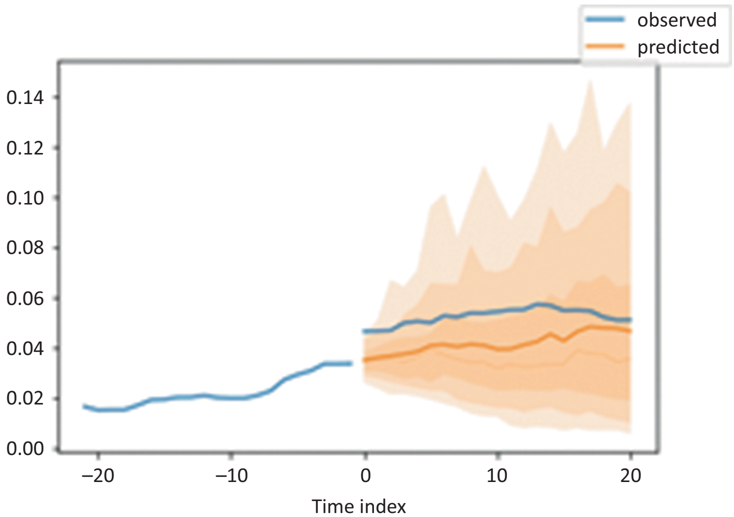


Figure 15.14: Volatility forecasts February 2022 DeepAR. (Compiled by authors.)

15.8 Conclusion

We find that DeepAR outperforms LSTM in all specifications and is well suited to the forecasting of Brent price volatility. It also yields accurate results during volatility shocks, which shows the model's versatility and ability to learn from various points in training data. The forecasting accuracy is astonishing, considering the only input variable is lagged data. We believe that our results could be improved with multivariate analysis. The DeepAR algorithm has been developed specifically to deal with multiple time series. As such, univariate forecasting does not exploit its features fully. A future addition to the model could therefore be the inclusion of control variables, such as different oil products, macroeconomic and volatility indices as well as parameters such as days of the week, month and year to allow the model to learn seasonal trends. It would also be interesting to analyse how the model deals with variables that follow a trend instead of a mean reverting series. DeepAR theoretically allows for different specifications of the probability density function to fit the statistical properties of the data. An analysis of the effects of function selection on algorithm accuracy could be worthwhile. While we find that DeepAR performs exceptionally, in the interest of further investigating which algorithms are well suited to volatility forecasting, additional algorithms could be included in our analysis. A comparison with GARCH models could be meaningful as could a hybridisation with GARCH input parameters.

Bibliography

Andersen, T. G., Bollerslev, T., Diebold, F. X., Ebens, H., Backus, D., Brandt, M., Deo, R., Engle, R., Granger, C., Hansen, L., Hasbrouck, J., Hentschel, L., Hurvich, C., de Lima, P., Schwert, B., Stambaugh, R., Tauchen, G., & Taylor, S. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, *61*(1), 43–76. https://doi.org/10.1016/S0304-405X(01)00055-1.

Baillie, R. T., Boierslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *74*(1), 3–30. https://doi.org/10.1016/S0304-4076(95)01749-6.

Barbaglia, L., Consoli, S., & Wang, S. (2021). Financial forecasting with word embeddings extracted from news: A preliminary analysis. In: Filipe, J., Ghosh, A. Oliveira Prates, E. & Zhou, L. (Eds.), Communications in Computer and Information Science: Vol. 1525. *Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (pp. 179–188). Cham: Springer. https://doi.org/10.1007/978-3-030-93733-1_12.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1.

Bourghelle, D., & Jawadi, F., & Rozin, P. (2021). Oil price volatility in the context of Covid-19. *International Economics*, *167*, 39–49. https://doi.org/10.1016/j.inteco.2021.05.001.

Carpio, L. (2019). The effects of oil price volatility on ethanol, gasoline, and sugar price forecasts. *Energy*, *181*(15), 1012–1022. https://doi.org/10.1016/j.energy.2019.05.067.

Chatziantoniou, I., Filippidis, M., Filis, G., & Gabauer, D. (2021). A closer look into the global determinants of oil price volatility. *Energy Economics*, *95*, 105092. https://doi.org/10.1016/j.eneco.2020.105092.

Cheng, Y., Yi, J., Yang, X., Lai, K., & Seco, L. (2022). A CEEMD-ARIMA-SVM model with structural breaks to forecast the crude oil prices linked with extreme events. *Soft Computing*, *26*, 8537–8551. https://doi-org.sire.ub.edu/10.1007/s00500-022-07276-5.

Chiroma, H., Abdul-Kareem,S., Shukri Mohd Noor, A., Abubakar, A.I., Sohrabi Safa, N., Shuib,L., Fatihu Hamza, M., Ya'u Gital, A., & Herawan, T. (2016). A review on artificial intelligence methodologies for the forecasting of crude oil price. *Intelligent Automation & Soft Computing*, *22*(3), 449–462. https://doi.org/10.1080/10798587.2015.1092338.

Consoli, S., Negri, M., Tebbifakhr, A., Tosetti, E., & Turchi, M. (2022). Forecasting the IBEX-35 stock index using deep learning and news emotions. In: Nicosia, G., Ojha, V., La Malfa, E., La Malfa, G., Jansen, G., Pardalos, P.M., Giuffrida, G. & Umeton, R. (Eds.), Lecture Notes in Computer Science: Vol 13163. *Machine Learning, Optimization, and Data Science*. Cham: Springer. https://doi.org/10.1007/978-3-030-95467-3_23.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196. https://doi.org/10.1093/jjfinec/nbp001.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1007. https://doi.org/10.2307/1912773.

Engle, R. F., & Patton, A. J. (2001). What good is a volatility model?. *Quantitative Finance*, 1(2), 237–245.

Ge, W., Lalbakhsh, P., Isai, L., Lenskiy, A., & Suominen, H. (2022). Neural network–based financial volatility forecasting: A systematic review. *ACM Computing Surveys*, *55*(1), 1–30. https://doi.org/10.1145/3483596.

Gupta, R., & Pierdzioch, C. (2022). Forecasting the realized variance of oil-price returns: a disaggregated analysis of the role of uncertainty and geopolitical risk. *Environmental Science and Pollution Research*, *29*(34), 52070–52082. https://doi-org.sire.ub.edu/10.1007/s11356-022-19152-8.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, *6*(2), 327–343.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Hull, J., & White, A. (1987). The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, *42*(2), 281–300. https://doi.org/10.1111/j.1540-6261.1987.tb02568.x

Jiao, X., Song, Y., Kong, Y., & Tang, X., 2022. Volatility forecasting for crude oil based on text information and deep learning PSO-LSTM model. *Journal of Forecasting*, *41*(5), 933–944.

Kamdem, J. S., Essomba, R. B., & Berinyuy, J. N. (2020). Deep learning models for forecasting and analyzing the implications of COVID-19 spread on some commodities markets volatilities. *Chaos, Solitons and Fractals*, *140*, 110215. https://doi.org/10.1016/j.chaos.2020.110215.

Kristjanpoller, W., & Hernández P, E. (2017). Volatility of main metals forecasted by a hybrid ANN-GARCH model with regressors. *Expert Systems with Applications*, *84*, 290–300. https://doi.org/10.1016/j.eswa.2017.05.024.

Kristjanpoller, W., & Minutolo, M. (2016). Forecasting volatility of oil price using an artificial neural network-GARCH model. *Expert Systems with Applications*, *65*, 233–241. https://doi.org/10.1016/j.eswa.2016.08.045.

Lang, K., & Auer, B. R., (2020). The economic and financial properties of crude oil: A review. *The North American Journal of Economics and Finance*, *52*(C), S1062940818302559. https://EconPapers.repec.org/RePEc:eee:ecofin:v:52:y:2020:i:c:s1062940818302559.

Li, Y., Jiang, S., Li, X. & Wang, S., (2021) The role of news sentiment in oil futures returns and volatility forecasting: Data-decomposition based deep learning approach. *Energy Economics*, *95*(C). https://ideas.repec.org/a/eee/eneeco/v95y2021ics0140988321000451.html.

Lu, X., Ma, F., Xu, J., & Zhang, Z. (2022). Oil futures volatility predictability: New evidence based on machine learning models. *International Review of Financial Analysis*, *83*, 102299. https://doi.org/10.1016/j.irfa.2022.102299.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*(2), 347–370. https://doi.org/10.2307/2938260.

Niu, H., & Zhao, Y. (2021). Crude oil prices and volatility prediction by a hybrid model based on kernel extreme learning machine. *Mathematical Biosciences and Engineering*, *18*(6), 8096–8122. https://doi.org/10.3934/mbe.2021402.

Poon, S.-H., & Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, *41*(2), 478–539. https://doi.org/10.1257/002205103765762743.

Ramos-Pérez, E., Alonso-González, P. J., & Núñez-Velázquez, J. J. (2019). Forecasting volatility with a stacked model based on a hybridized Artificial Neural Network. *Expert Systems with Applications*, *129*, 1–9. https://doi.org/10.1016/j.eswa.2019.03.046.

Rakpho, P., Yamaka, W., & Phadkantha, R. (2022). Predicting energy price volatility using hybrid artificial neural networks with GARCH-type models. In: Honda, K., Entani, T., Ubukata, S., Huynh, V.-N., & Inuiguchi, M. (Eds.), Lecture Notes in Computer Science: Vol. 13199. *Integrated Uncertainty in Knowledge Modelling and Decision Making*. Cham: Springer. https://doi.org/10.1007/978-3-030-98018-4_26.

Ribeiro, G. T., Alves Portela Santos, A., Cocco Mariani, V., & dos Santos Coelho, L. (2021). Novel hybrid model based on echo state neural network applied to the prediction of stock price return volatility. *Expert Systems with Applications*, 184. https://doi.org/10.1016/j.eswa.2021.115490.

Salinas, D., Flunkert, V., & Gasthaus, J. (2017). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. http://arxiv.org/abs/1704.04110.

Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, *36*(3), 1181–1191. https://doi.org/10.1016/j.ijforecast.2019.07.001.

Sehgal, N., & Pandey, K.K. (2015). Artificial intelligence methods for oil price forecasting: A review and evaluation. *Energy Systems*, *6*(4), 479–506. https://doi.org/10.1007/s12667-015-0151-y.

Sentana, E. (1995). Quadratic ARCH models. *The Review of Economic Studies*, *62*(4), 639–661. https://doi.org/10.2307/2298081.

Tissaoui, K., Zaghdoudi, T., Hakimi, A., Ben-Salha, O., & Ben Amor, L. (2022). Does uncertainty forecast crude oil volatility before and during the COVID-19 outbreak? Fresh evidence using machine learning models. *Energies*, *15*(15), 5744. https://doi.org/10.3390/en15155744.

Verma, S. (2021). Forecasting volatility of crude oil futures using a GARCH–RNN hybrid approach. *Intelligent Systems in Accounting, Finance & Management*, *28*(2), 130–142. https://doi.org/10.1002/isaf.1489.

Vidal, A., & Kristjanpoller, W. (2020). Gold volatility prediction using a CNN-LSTM approach. *Expert Systems with Applications*, 157. https://doi.org/10.1016/j.eswa.2020.113481.

Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, *18*(5), 931–955. https://doi.org/10.1016/0165-1889(94)90039-6.

Zhang, Y., Ma, F., & Wei, Y. (2019). Out-of-sample prediction of the oil futures market volatility: A comparison of new and traditional combination approaches. *Energy Economics*, *81*, 1109–1120. https://doi.org/10.1016/j.eneco.2019.05.018.

Zhang, Y., Wahab, M., & Wang, Y. (2023). Forecasting crude oil market volatility using variable selection and common factor. *International Journal of Forecasting*, *39*(1), 486–502. https://doi.org/10.1016/j.ijforecast.2021.12.013.

Zhang, L., Wang, J, & Wang, B. (2020). Energy market prediction with novel long short-term memory network: Case study of energy futures index volatility. *Energy*, *211*, 118634. https://doi.org/10.1016/j.energy.2020.118634.

Jan Vogt, Alexander Bönner, Michael Römmich, Malte Weiß,
Merih Türkoglu

Chapter 16
Energy Stock Price Forecast Based on Machine Learning and Sentiment Analysis – Which Approach Performs Best in Day Trading?

Abstract: We explore the application of machine learning (ML) methods to predict energy stock prices. We apply established ML methods, Gradient Boosted Regression Trees (GBRT), Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) trained on the energy market stock data retrieved from Uniper from January 2019 to August 2020. Furthermore, we incorporate sentiment data linked to Twitter contributions in the aforementioned time period. We apply these algorithms to predict the next day's close price for three Energy companies (Uniper, Enel, EDF), and simulate a buy-and-hold trading strategy to measure the performance of our models. Our results indicate that MLP yields the most accurate predictions with the lowest mean absolute error. Applied to Uniper stock market data, our trading simulation significantly outperforms the buy-and-hold benchmark. Furthermore, the results show that the use of sentiment values improves trading performance significantly.

Keywords: energy market, stock price prediction, sentiment analysis, LSTM, GBRT, MLP

16.1 Introduction

In recent years, machine learning (ML) algorithms have been applied to various fields, such as physical simulations, weather forecasts, programming and language generation. In particular, stock market predictions represent a lucrative application that has gained significant attention (Islam, Hasan & Khan 2021; Liu, Dang & Yu 2020;

Jan Vogt, sipgate GmbH, Düsseldorf, e-mail: contact@j-vogt.com
Alexander Bönner, Hochschule Ruhr West, Mülheim a.d. Ruhr,
e-mail: alexander.boenner@hs-ruhrwest.de
Michael Römmich, Hochschule Ruhr West, Mülheim a.d. Ruhr,
e-mail: michael.roemmich@hs-ruhrwest.de
Malte Weiß, Hochschule Ruhr West, Bottrop, e-mail: malte.weiss@hs-ruhrwest.de
Merih Türkoglu, Hochschule Ruhr West, Bottrop, e-mail: merih.tuerkoglu@hs-ruhrwest.de

https://doi.org/10.1515/9783110790320-016

Wang & Guo 2020). While various market models have been developed using ML algorithms, there is little evidence of the use of ML algorithms in the energy stock market.

In this chapter, we explore three ML methods to predict stock prices in the energy market. We employ Gradient Boosted Regression Trees (GBRT), Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) to create predictive models for three energy companies (i.e., Uniper, Enel and EDF). Our input vector consists of two data sets. The first data set constitutes stock prices (high, low, open and close) of Uniper in the period from January 2019 to August 2020. The second data set incorporates sentiment data extracted from Twitter publications of Uniper. We analyse the accuracy of stock predictions with and without sentiment data, and evaluate our model using a buy-and -hold strategy as a benchmark.

16.2 Related Work

16.2.1 Machine Learning and Sentiment Analysis Applied on Stock Price Predictions

Achkar et al. (2018) used an MLP and the LSTM to predict the daily closing prices of Alphabet, Facebook and Bitcoin shares. Their analysis indicates that the deviations measured as mean squared error (MSE) used to predict the closing price ranged from 3% to 16% for the MLP model and from 0.5% to 12% for the LSTM model, respectively. The authors find that neural networks can be used as an effective and promising tool for stock market predictions.

Torres et al. (2019) used an MLP and a decision tree to make predictions for the Apple stock. Historical price data (open, close, high, low and trading volume) over the last 250 trading days built the foundation of their analysis. The utilised methodology aimed to predict the closing price of a share on a particular day and, also, shows tolerable errors mean absolute error (MAE) results. The authors conclude that both methods are suitable for predicting stock prices and note that sentiment analysis could further enhance the accuracy of predictions.

Kolasani and Assaf (2020) investigated the effectiveness of using tweets to predict stock prices. Moreover, the authors looked at the effectiveness of neural networks in predicting stock price movements in contrast to traditional ML. The study elaborates on a study by Chakraborty et al. (2017) to predict price movements of the stock of the Apple and the Dow Jones index. Both studies examined several models, such as Logistic Regression, Support Vector Machine, Decision Tree, GBRT and Random Forests, to represent the manually tested sentiments in an ML model. Kolasani and Assaf (2020) found that, on average, the MLP performs better in terms of MAE and root mean squared error (RSME) in predicting the price difference of stocks than the GBRT. The authors note that the GBRT tends to be a more optimistic approach, while the MLP tends to be more pessimistic in predicting a price difference.

Hutto and Gilbert (2014) implement a lightweight and rule-based model for general sentiment analysis called Valence Aware Dictionary for sEntiment Reasoning (VADER). It can be applied to social media style text and multiple domains such as news services or ratings. The model includes a generalisable, valence-based, human-organised lexicon for sentiment. VADER provides four values "positive," "negative," "neutral" and "composite" to identify sentiment. The attribute "composite" describes a composite value, namely the sum of the valence values of each word in the lexicon, which is calculated and then normalised to a value between –1 (negative) and +1 (positive) (Hutto & Gilbert, 2014).

Deepika and Nirupamabhat (2020) use different models to derive price direction prediction using stock data (open, close, adjusted close, high, low and trading volume) of Apple, Amazon, Infosys, Microsoft, Oracle and TCS. This was combined with technical analysis and sentiment analysis. An optimised Artificial Bee Colony (ABC)-LSTM proves to be the best model in terms of mean absolute percentage error (MAPE) compared to also optimised Least-squares Support Vector Machine (LSSVM), Gradient Boost, LSTM, ABC-LSSVM and ABC-Gradient Boost.

16.2.2 Machine Learning and Trading Strategy

In order to put ML approaches into trading practices, these forecasting approaches are trained and thus optimised with respect to metrics like the MSE, MAE or alike. However, from a trader's perspective, it is important that forecasting models perform well in monetary terms like profits or Sharpe Ratios. The following research articles deal with the performance of translating ML strategies into trading results.

Mittermayer (2004) forecasted intraday stock price trends for a short time immediately after the publication of press releases by his NewsCATS system, which used "Support Vector Machines" (SVM) for ML. Categorised press releases built the basis for trading strategies. The author showed that categorised press releases contain additional information in such a way that NewsCATS short-term trading strategy outperformed trading strategies by randomly buying or shorting stocks directly after the publication of press releases.

Fischer & Krauss (2018) used LSTM networks to predict out-of-sample directional stock movements for S&P 500 stocks only based on constituent S&P 500 stock data from 1992 to 2015. Their approach outperforms by daily returns and by Sharpe Ratio memory-free classification methods like random forests or logistic regression classifiers. Furthermore, the researchers successfully translated the output of the model (i.e., the selected stocks) into a rule-based short-term trading strategy to partially lighten up the black box character of the LSTM network.

Yuan et al. (2020) focussed on the Chinese stock market from 2010 until 2018 and incorporated different methods, including ML applied to long/short-term trading strategies. Their predictions on the stock price trend were based on three approaches: Support Vector Machine (SVM), Random Forests (RF) and an Artificial Neural Network (ANN). Long/short-trading strategies were evaluated by indicators like annualised returns, Sharpe Ratios or profit-loss ratios, etc. Overall, the RF approach performed best in their analysis.

Li et al. (2021) applied tensor theory to present fundamental stock information (high/low prices, volume, turnover, P/E-Ratio, P/B-Ratio, etc.) with media information (positive, negative or divergent media sentiment) to preserve the multifaceted and interrelated nature of both streams of information. An LSTM model is proposed to capture the relations between market information and stock movements. Based on the Sharpe Ratio reflecting the trade-off between risk and return, their approach outperformed trading algorithms in an investment simulation of the Chinese stock market.

16.3 Methodology and Data

16.3.1 Data Procurement and Preparation

Based on the findings from the literature review, we structure the methodology of our study as follows. In the first step, the data set is generated and forms the basis for training neural networks and decision trees to predict future stock price trends. Company profiles of Europe's three major energy groups, Enel, EDF and Uniper are identified. Data used to predict price movements consists of historical share price developments and user tweets retrieved from Twitter (which is used for sentiment analysis). The data was collected for the period from January 2019 to December 2020. Historical stock price movements were obtained from Yahoo Finance using the Python library, yfinance. For each of the three groups, information was collected over 512 trading days. Opening, closing, high and low prices were recorded for each trading day.

The Twitter REST API V2, which is reserved for non-commercial projects, was used to determine user contributions. The data collection was carried out in August 2021. The data collection included the number of tweets found on the public profiles of the three listed companies as can be seen in Table 16.1. Whenever company profiles were mentioned, a correlation between shares and posts on social media was established. For this purpose, all official profiles for each company were identified in advance.

The data was pre-processed for sentiment analysis. Twitter hashtags come in various forms, for example, as individual words such as nouns and adjectives, and as groups of words that are combined primarily with the camel or Pascal spelling. We normalised these hashtags by breaking them down into individual words. Finally, parts in tweets that would not contribute to sentiment analysis were removed before-

Table 16.1: Companies Twitter-Profile with the corresponding numbers of Tweets during the research period from January 2018 to December 2020. (Compiled by authors.)

Company	Twitter-Profile	Numbers of tweets
UNIPER	@uniper_energy	37,179
ENEL	@Enelgroup, @EnelGroup	51,500
EDF	@edfenergy	53,840
		Total: 142,519

hand. These included URLs or evaluative usernames that were used whenever Twitter was mentioned. In addition, up to 43 different languages were identified with Google Language Detection. Following this, the multilingual posts were translated into English as the common target language.

16.3.2 Sentiment Analysis and Data Set Building

In this study, VADER analysis is used to extract sentiments from Twitter posts (Hutto & Gilbert, 2014). VADER was developed specifically for social media vocabulary and is therefore also suitable for Twitter posts. VADER is based on an English vocabulary. Consequently, it is required that messages are exclusively written in English. The process of sentiment analysis is as follows. Firstly, the so-called polarity of the sentiment of each tweet is determined sequentially for each company. Secondly, sentiment values were determined for the tweets. The mean values were calculated as overall sentiment values for a trading day. The sentiment of a day is thus defined by the average mean value of the sentiment for each day. The influence of these attributes could be investigated in later studies. When creating the data set, mean values returned from the sentiment analysis as well as the share price movements for each trading day were combined and formed the input vector for the ML models. The closing price of the following day was used as the target value in this study (t + 1).

16.3.3 Model Development

16.3.3.1 Hyperparameter Setting

Two versions of each of the models are trained, one using historical stock price movements to predict the next day's closing price and the other including sentiment data to investigate the impact. The split is 80:20, with 80% of the data used as training data and 20% of the data used as test data. The training data is used to train the model, while the test data is used for evaluation. For each model, multiple metrics are collected for evaluation. The metrics are primarily used to evaluate the accuracy and secondly to benchmark the two models. The metrics collected are the error values in the form of MSE, RMSE and MAE. In this study, as depicted in Table 16.2, three different levels of complexity are examined for each of the models (GBRT, MLP and LSTM).

Table 16.2: Complexity table. Displaying complexity of each model. (Compiled by authors.)

Complexity	GBRT	MLP	LSTM
Low	Estimator: 1.000 Max Depth: 4	Layers: 4-2-1	Layers: 4-2-1
Medium	Estimator: 1.000 Max Depth: 8	Layers: 8-4-1	Layers: 8-4-1
High	Estimator: 1.000 Max Depth: 16	Layers: 16-8-1	Layers: 16-8-1

In the course of the model developments, a series of measurements of different hyperparameters are carried out in order to compare them with each other and to find a suitable configuration. Different constellations of hyperparameters are tested and the resulting error values from the measurement series are compared with each other. Dropout (0%, 10%, 20%), batch (1, 2, 4) and the learning rate of (0.1, 0.01, 0.001) are considered. Stochastic Gradient Descent (SGD) is used as the optimiser to minimise the neural network errors. Furthermore, the MSE is chosen for the error function.

16.3.3.2 Hyperparameter Tuning with Uniper

Uniper is chosen to determine the hyperparameters. We hypothesise that the configuration of one model leads to similar results for other energy groups. The series of measurements are carried out sequentially, with one parameter being investigated without including another. GBRT, LSTM and MLP require a learning rate as a basis for optimising the model, so we start with this hyperparameter and then continue with batch size and dropout. The investigation is applied to all complexity levels “low,” “medium” and “high.” The MSE is used as a metric to evaluate the investigation.

The MLP results showed an increased error value at a learning rate of 0.1 across all complexity levels. The learning rate of 0.001 resulted in the lowest error values. No large deviations were found in the measurement series for the LSTM but led to the same values as the MLP at a learning rate of 0.001. The error values of the GBRT showed that a learning rate of 0.01 led to the lowest error values at all complexity levels, both for the data set with historical stock values and for the included sentiment with identical hyperparameter settings. This concludes the investigation for the GBRT, as only the learning rate was considered as a hyperparameter. After the learning rate leads to the lowest values in all complexity levels, the consideration of the lot sizes follows. A comparison between 1, 2 and 4 is carried out and applied to the learning rate of 0.001 determined previously.

In the series of experiments of the MLP, batch size 2 and 4 showed a lower error value in contrast to batch size 1. Among them, batch size 4 gave the lowest error values for MSE Stock and MSE Sentiment in all complexity levels. The batch size results for LSTM differ from the previously collected data. For the complexity level "simple" with the structure 4-2-1, lot size 2 leads to the lowest error values. For the complexity level "medium" and "complex," the lowest error values were measured for lot size 4. For the following consideration of the LSTM of the dropout, different batch sizes are therefore used for the complexity levels.

After the learning rate and the batch size were determined, the last missing hyperparameter, dropout, was determined. Here, the hyperparameters that resulted in the lowest error values in the previously determined tests were used. The results of the MLP as well as the LSTM show that in all complexity levels, the use of dropouts leads to higher error values for both MSE-Stock and MSE-Sentiment.

As a result, for the remainder of the study, the GBRT is used with a depth of 4 and the estimator of 1000 and a learning rate of 0.01. The MLP is used with the hyperparameters dropout 0%, LR 0.001, batch size 4 and the model setup of 16-8-1. The LSTM gave the same results as the MLP but with the difference that the model setup of 8-4-1 is used in this case.

16.4 Results

We use error functions as metrics like the L2-Error to mathematically validate the outcomes of each model for all companies. Then, using a variety of day trading methods based on the model, we test if the mathematical results correspond to actual world outcomes.

16.4.1 Error Scores and Analysis

The historical stock model and the sentiment model from 17.3 were individually applied to MSE, RMSE and MAE metrics, respectively. With the use of this analysis, we can determine whether the target value and prediction have any relationship to stock or sentiment values. The models have been trained on Uniper's data and tested on three companies: Uniper, Enel and EDF. The input for the stock models included open, low, high and close values of the prior trading day. For the sentiment model, the same values along with the mean sentiment on the previous trading day were employed. The evaluation period was between 14 August 2020 and 29 December 2020 for each company with a total of 100 trading days. The following Table 16.3 displays the findings for each company and model assessed using the MSE. For more measurement results, please see the appendix.

Table 16.3: Analysis on each model with MSE. (Compiled by Authors.)

ML-Model	Model	Error	Uniper	Enel	EDF
GBRT	STOCK	MSE	0.1404 . . .	0.0195 . . .	0.0768 . . .
GBRT	SENTIMENT	MSE	0.1410 . . .	0.0228 . . .	0.0719 . . .
MLP	STOCK	MSE	0.1305 . . .	0.0151 . . .	0.0564 . . .
MLP	SENTIMENT	MSE	0.1287 . . .	0.0179 . . .	0.0570 . . .
LSTM	STOCK	MSE	0.1455 . . .	0.0468 . . .	0.0825 . . .
LSTM	SENTIMENT	MSE	0.1422 . . .	0.0405 . . .	0.0770 . . .

16.4.1.1 Uniper

The analysis shows that the MSE of Uniper GBRT using the stock model is 0.1404 US dollar (USD), MLP 0.1305 USD and LSTM 0.1455 USD, while the sentiment model GBRT is 0.1410 USD, MLP 0.1287 USD and LSTM 0.1422 USD (see Table 16.3 and Appendix). MLP and LSTM show a slightly better performance with sentiment, while GBRT performs worse. Interestingly, Uniper's analysis results have higher error values for MSE, RMSE and MAE compared to the other two companies. The probable reason for this is that Uniper's share value per stock is nearly twice of Enel and EDF. Figure 16.1 displays the 100-day evaluation period and the respective prediction of both MLP models. Stock and sentiment models have a small difference of 0.0018 USD on the analysis, but the figure exposes larger differences across all trading days.

16.4.1.2 Enel

As for Enel the GBRT value is 0.0195 USD, MLP 0.0151 USD and LSTM 0.0468 USD. While for the sentiment model, the GBRT is 0.0228 USD, MLP 0.0179 USD and LSTM 0.0405 USD.

In contrast to Uniper, only LSTM gained an improvement in performance by using sentiment values, while GBRT and MLP displayed a worsening in performance. Figure 16.2 shows the MLP predictions of Enel. Like Uniper, the projections are pessimistic, which means that most of them are below the target value. On some dates, the real values also have a small visual delay, for example between 1 November 2020 and 15 November 2020.

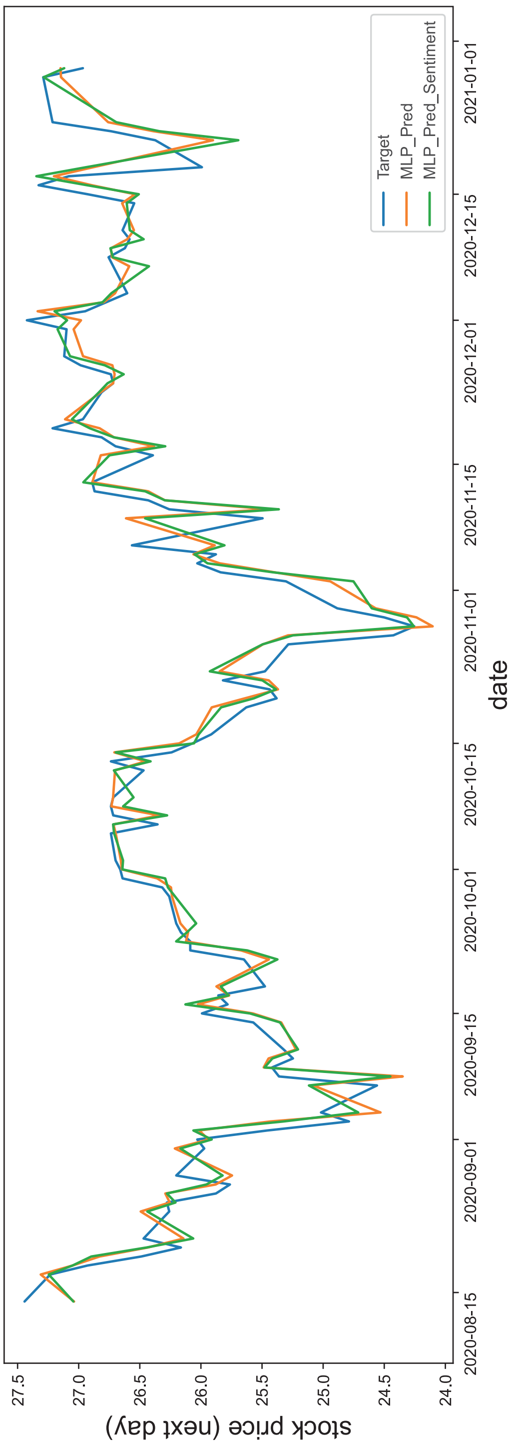


Figure 16.1: MLP validation on Uniper's 100 trading days. Predictions of stock and sentiment MLP. Target is the real value for the time period, while orange shows stock model prediction and green displays sentiment model. Sentiment and stock show strong similarities. (Compiled by authors.)

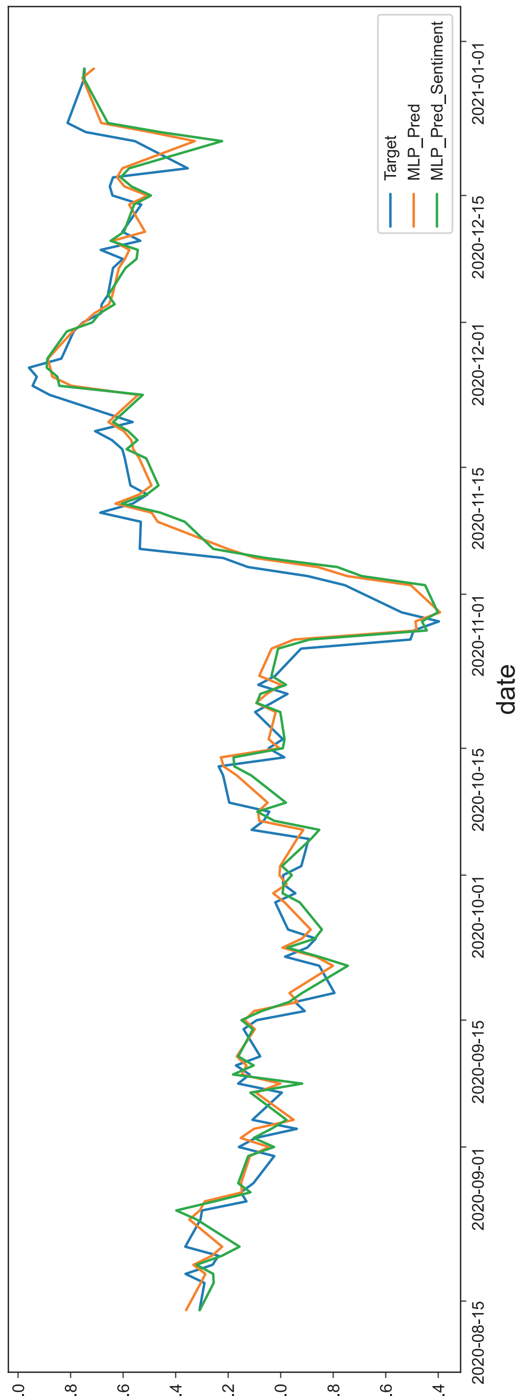


Figure 16.2: MLP Validation on Enel. (Compiled by authors.)

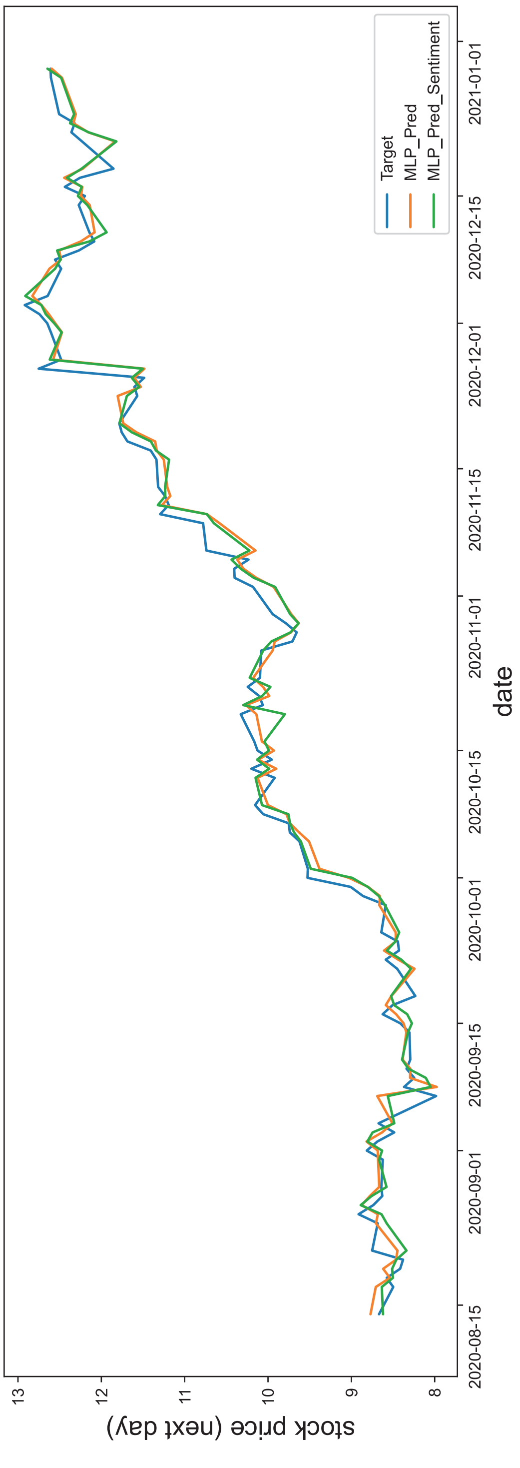


Figure 16.3: MLP Validation on EDF. (Compiled by authors.)

16.4.1.3 EDF

For EDF, the GBRT stock model produced values of 0.0768 USD, 0.0564 USD for MLP and 0.0825 USD for LSTM, while the sentiment model result in values of 0.0719 USD for GBRT, 0.0570 USD for MLP and 0.0770 USD for LSTM, respectively. The MLP's prediction for the EDF is shown in Figure 16.3. Compared to Uniper and Enel, the sentiment and stock prices seem to be more in line here.

Given that the sentiment has somewhat improved performance on the MSE, except for MLP, the results for the MSE on EDF are completely different from those for Uniper and Enel. MSE performed somewhat worse with sentiment for the MLP. Overall, all models for EDF have slightly improved sentiment performance.

In our study of the three distinct companies, MLP shows the best performance and the lowest error rates across all three metrics. Interestingly, the LSTM produced the highest error for the three companies.

16.4.2 Applied Trading Strategy

16.4.2.1 Derivation and Implementation of a Straightforward Trading Strategy

In the following, a trading strategy is derived for the stocks of Uniper, Enel and EDF. The basis for the trading strategy are the forecast values of the model with the most robust performance, the MLP with layers 16-8-1, a learning rate of 0.001, a batch size 4 and a dropout 0. Both variants, that is, the forecast values with sentiment analysis and without sentiment analysis, are examined for the test data period.

The aim is to achieve the best possible return compared to a buy-and-hold strategy (benchmark) for the next 100 trading days. The investable amount of money at day 1 is 100 EUR. Following a signal, it is possible to buy the shares of the respective company at the open price at the beginning of a trading day and automatically sell them at the close price at the end of the trading day. On the other hand, it is also possible, after receiving a respective signal, to sell shares at the value of the investable amount of money at the open price at the beginning of a trading day and automatically buy them back at the close price at the end of the trading day. The resulting amount is assumed to be investable (long or short strategy accordingly) again on the next day. Borrowing fees and transaction costs are not further considered in the analysis. The investment in fractional shares is possible.

In the next step, buy and sell signals are determined. For this purpose, the forecast value for the day, measured at the end of the previous day, is compared to the close price of the previous day. If the forecast value is greater than the close price, shares at the value of the investable amount are purchased at the open price on the following day. In the opposite case, shorting shares at the value of the investable amount at the open price is undertaken.[1]

16.4.2.2 Results and Interpretation

After the 100 trading days of the test data period, the total returns for all three companies are as shown in Table 16.4.

Table 16.4: Total returns of the trading strategy compared with benchmark. (Compiled by authors.)

	Return with sentiment	Return without sentiment	Return benchmark
Uniper	30.28%	15.03%	−6.66%
Enel	10.91%	5.09%	5.56%
EDF	5.12%	−18.06%	46.74%

On the one hand, it can be seen that the model with input of the sentiment analysis leads to significantly better results for all three companies than without input of the sentiment analysis. Moreover, the model with input of the sentiment analysis leads to positive returns for all three companies. According to an applied t-test, Uniper shows a significant outperformance compared to the buy-and-hold strategy (benchmark). Enel´s result is positive compared to the buy-and-hold strategy as well, but not on a significant level. In contrast, the returns forecasted for EDF are outweighed by the buy-and-hold strategy. The same pattern is valid for the respective Sharpe ratios.

A possible explanation for the different performance of the trading strategy compared to the buy-and-hold strategy could be that the share prices of Uniper and Enel tended to move sideways (Uniper slightly down, Enel slightly up) during the test period. On the other hand, the share price of EDF has developed strongly and positively with a significantly higher volatility in share price than Uniper and Enel. This could be interpreted in such a way that the trading strategy works better in less volatile markets without a clear trend but seems less suitable in nervous and strongly trending markets. However, it should be noted that EDF also generated a positive return in the chosen period.

1 The strategy was further refined, including the inclusion of volatility or the previous day's forecast quality. However, there were no significant improvements in the results.

16.5 Limitations and Future Work

The present analysis is subject to various restrictions. For example, it is limited to three levels of complexity in the construction of the models. Furthermore, the subsequent calibration of the models, including hyperparameter setting, focuses on only one company. Both restrictions are due to the available computer capacities. A detailed investigation of the hyperparameters could lead to different and possibly better results. Other ML methods, such as CNN, could also be applied in principle.

The sentiment analysis was based on Twitter tweets about the respective companies in all languages used. Since free products did not have sufficient translation quality, Google Translate was used. Thus, the sentiment analysis in this study could only be carried out for three companies and especially tweet leader RWE could not be further investigated due to these restrictions.

Furthermore, additional factors such as likes, number of positive and negative comments or the evaluation of the content behind the links used could provide further insights. Likewise, alternative news portals can contribute further useful information on price changes. On top of this, the influence of different time intervals on the sentiment values could be looked at, which include several days or a lag, for example.

In addition, a follow-up study could use tools that are more specialised in sentiment analysis in the financial sector, such as FinBERT, instead of VADER. Furthermore, the study was limited to a period of 2 years due to capacity bottlenecks in the processing of sentiment inputs. A longer period could lead to lower error values.

Especially interesting is the pattern where the error values for the MLP model of Enel and EDF are slightly lower for the stock models whereas the trading strategy performs better for the sentiment models. The signals of the sentiment models seem to be more significant. Why this is the case should be investigated in the next step.

The analysis presented could be made even more sophisticated, especially in the trading section, if the data set of share prices, which is limited to open, close, high and low, was extended to include, for example, shorter time intervals such as minute-by-minute price movements.

For the purposes of finding out whether the hyperparameters can also be applied to other price movements in the energy sector, we transfer the hyperparameters used in the study to the 48 largest energy companies. We use the same time period for this purpose and measure the error value with the MSE. The results based on historical stock price movements are promising. The hyperparameters for ML models yielded the lowest error values for 22 GBRT, 6 MLP and 4 LSTM companies. Looking at the first quartiles according to the hyperparameters obtained for Uniper, of the 48 possibilities, 44 could be counted for GBRT (91.66%), 45 for MLP (93.75%) and 41 for LSTM (85.41%). These results encourage us to perform further studies in this area in the future.

16.6 Conclusion

Exploring the application of various ML techniques showed that sentiment analysis adds value to day-trading strategies to predict stock prices in energy markets. Generally, our results indicate that MLP yields the most accurate predictions with lowest mean absolute error. Applied to Uniper stock, our trading simulation significantly outperforms a 100-day buy-and-hold strategy as a benchmark.

Appendix

Tables 16.5–16.7 show the results of the analysis on each company.

Table 16.5: Results of the different metrics used for each model trained on Uniper. (Compiled by authors.)

ML-Model	Model	MSE	RMSE	MAE
GBRT	STOCK	0.1404 . . .	0.3747 . . .	0.2892 . . .
GBRT	SENTIMENT	0.1410 . . .	0.3756 . . .	0.2897 . . .
MLP	STOCK	0.1305 . . .	0.3612 . . .	0.2751 . . .
MLP	SENTIMENT	0.1287 . . .	0.3586 . . .	0.2569 . . .
LSTM	STOCK	0.1455 . . .	0.3814 . . .	0.2979 . . .
LSTM	SENTIMENT	0.1422 . . .	0.3771 . . .	0.2851 . . .

Table 16.6: Results of the different metrics used for each model trained on Enel. (Compiled by authors.)

ML-Model	Model	MSE	RMSE	MAE
GBRT	STOCK	0.0195 . . .	0.1398 . . .	0.1100 . . .
GBRT	SENTIMENT	0.0228 . . .	0.1509 . . .	0.1152 . . .
MLP	STOCK	0.0151 . . .	0.1229 . . .	0.0931 . . .
MLP	SENTIMENT	0.0179 . . .	0.1337 . . .	0.1017 . . .
LSTM	STOCK	0.0468 . . .	0.2164 . . .	0.1743 . . .
LSTM	SENTIMENT	0.0405 . . .	0.2013 . . .	0.1615 . . .

Table 16.7: Results of the different metrics used for each model trained on EDF. (Compiled by authors.)

ML-Model	Model	MSE	RMSE	MAE
GBRT	STOCK	0.0768 . . .	0.2772 . . .	0.1979 . . .
GBRT	SENTIMENT	0.0719 . . .	0.2682 . . .	0.1971 . . .
MLP	STOCK	0.0564 . . .	0.2374 . . .	0.1725 . . .
MLP	SENTIMENT	0.0570 . . .	0.2387 . . .	0.1719 . . .
LSTM	STOCK	0.0825 . . .	0.2872 . . .	0.2148 . . .
LSTM	SENTIMENT	0.0770 . . .	0.2775 . . .	0.2108 . . .

Bibliography

Achkar, R., Elias-Sleiman, F., Ezzeddine, H., & Haidar, N. (2018). Comparison of BPA-MLP and LSTM-RNN for stocks prediction. *International Symposium on Computational and Business Intelligence (ISCBI)*, 48–51. https://doi.org/10.1109/ISCBI.2018.00019.

Chakraborty, P., Pria, U. S., Rony, M. R. A. H., & Majumdar, M. A. (2017). Predicting stock movement using sentiment analysis of Twitter feed. *International Conference on Informatics, Electronics and Vision*, 1–6. https://doi.org/10.1109/ICIEV.2017.8338584.

Deepika, N., & Nirupamabhat, M. (2020). An optimized machine learning model for stock trend anticipation. *IngéNierie des Systèmes D Information*, *25*(6), 783–792. https://doi.org/10.18280/isi.250608.

Fischer T., & Krauss C. (2018). Deep learning with long-short term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054.

Islam, S. B., Hasan, M. M., & Khan, M. M. (2021). Prediction of stock market using recurrent neural network. *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 479–483. https://doi.org/10.1109/IEMCON53756.2021.9623206.

Hutto, C. J., & Gillbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225. https://doi.org/10.1609/icwsm.v8i1.14550.

Kolasani, S. V., & Assaf, R. (2020). Predicting stock movement using sentiment analysis of Twitter feed with neural networks. J*ournal of Data Analysis and Information Processing*, *8*(4), 309–319. https://doi.org/10.4236/jdaip.2020.84018.

Li, Q., Tan, J., Wang, J., & Chen, H. (2021). A multimodal event-driven LSTM model for stock prediction using online news. *IEEE Transactions on Knowledge and Data Engineering*, *33*(10), 3323–3337. https://doi.org/10.1109/TKDE.2020.2968894.

Liu, Z., Dang, Z., & Yu, J. (2020). Stock price prediction model based on RBF-SVM algorithm. *2020 International Conference on Computer Engineering and Intelligent Control (ICCEIC)*, 124–127. https://doi.org/10.1109/ICCEIC51584.2020.00032.

Mittermayer, M.-A. (2004). Forecasting Intraday stock price trends with text mining techniques. *37th Annual Hawaii International Conference on System Sciences*. 10. https://doi.org/10.1109/HICSS.2004.1265201.

Torres P. E. P., Hernández-Álvarez, M., Hernández, E. A. T., & Yoo, S. G. (2019). Stock market data prediction using machine learning techniques. In: Á. Rocha, C. Ferrás, & M. Paredes (Eds.), Information Technology and Systems. *Proceedings of ICITS 2019* (pp. 539–547). Springer. https://doi.org/10.1007/978-3-030-11890-7_52.

Wang, Y., & Guo, Y. (2020). Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost. *China Communications*, *17*(3), 205–221. https://doi.org/10.23919/JCC.2020.03.017.

Yuan, X., Yuan, J., Jiang, T., & Ain, Q. U. (2020). integrated long-term stock selection models based on feature selection and machine learning algorithms for China Stock Market. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.2969293.

Gregor Johnen, Jens Kley-Holsteg, André Niemann, Florian Ziel

Chapter 17
Optimising Water Supply – Application of Probabilistic Deep Neural Networks to Forecast Water Demand in the Short Term

Abstract: Ensuring water supply-demand balance has become increasingly vital for supply companies due to climate change, demographic shifts and unforeseen future changes. Specifically, efficient management of water resources is necessary during prolonged dry spells and hotter summer days when demand for water peaks. By providing precise water demand forecasts, suppliers can proactively manage their facilities and prevent shortages or the depletion of the available sources.

Since water demand is influenced by various factors, including weather conditions and human activities, it is essential to distinguish their impacts. This study analyses the effect of weather variables through comprehensive statistical analysis, revealing intricate, multidimensional effects that are difficult to model. In addition to historical district-level data from West and Central Germany, we use historical weather variables and calendar data at an hourly resolution.

We present a non-linear deep learning model, compared against a Lasso-estimated time series model and a classical $AR(p)$ model, which is designed to operate across various lead times (24–240 hours). By prioritising performance, accuracy and interpretability, these models are well suited for real-time application and offer comprehensive probabilistic forecasts. The assessment of the model is carried out by utilising appropriate metrics for point and probabilistic forecasts. By assessing the predictors' impact, we measure the continuing effects of climate change on water usage.

Keywords: urban water demand, optimisation, forecasting, time series, uncertainty, decision-making, machine learning

Gregor Johnen, Institute of Hydraulic Engineering and Water Resources Management, University of Duisburg-Essen, Essen, e-mail: gregor.johnen@uni-due.de
Jens Kley-Holsteg, Chair of Environmental Economics esp. Economics of Renewable Energy, University of Duisburg-Essen, Essen
André Niemann, Institute of Hydraulic Engineering and Water Resources Management, University of Duisburg-Essen, Essen
Florian Ziel, Environmental Economics esp. Economics of Renewable Energy, University of Duisburg-Essen, Essen

https://doi.org/10.1515/9783110790320-017

17.1 Introduction

Forecasting water demand is important for managing the water supply-demand balance in a given region. Accurate forecasting allows water utilities and governments to address potential shortages in advance through making informed decisions about infrastructure development and resource allocation. Especially increasing peak demands in the summer months poses a growing challenge as these demands are approaching the maximum capacities of what water distribution networks can provide. A robust estimate of future water demand in the short term (24 hours to 1 week) is therefore required to allow water managers and policymakers to optimise drinking water distribution and maintain a proper balance between supply and demand under an uncertain future (de Souza, Costa & Libânio, 2019).

However, forecasting water demand can be difficult due to the complexity of the factors that influence consumption, such as population growth, economic conditions, changing weather patterns and changes in water usage behaviours (Xenochristou, 2019). Additionally, the quality and availability of data can be a challenge, as well as the need to consider different forecast horizons and spatial scales.

17.1.1 Forecasting Water Demand

Water demand forecasting has been subject to researchers and environmental engineers for many years now (de Souza, Costa & Libânio, 2019; Donkor et al., 2014; Ghalehkhondabi et al., 2017). Multiple forecasting techniques have been developed that can be used at different forecast horizons, from short term to long term.

Initially, linear regression and time series models were used, as outlined by Adamowski (2008), who compared multiple linear regression and time series analysis techniques for peak daily summer water demand forecast modelling. Innovations of machine learning (ML) algorithms have led to advances in the capability to predict water demand in the short, medium and long term as compared to the more classical approaches (Kley-Holsteg & Ziel, 2020; Tiwari & Adamowski, 2015; Xenochristou & Kapelan, 2020). Still, the more researched artificial neural network (ANN)-based models are among the most commonly used ML techniques to predict water demand as for example applied by Kühnert et al. (2021) or Guo et al. (2018) and are often proposed in the literature as best if properly tuned (Menapace, Zanfei & Righetti, 2021).

One drawback of such “black box” style of models is that they are harder to interpret as compared to linear models. After Xenochristou (2019), this implies that while they can achieve high accuracy, their results are rarely used directly for demand management strategy development and planning. One potential reason for this fact is that to date, very few examples have implemented advanced ML models in a probabilistic way as reviewed by Donkor et al. (2014). However, detailed information on forecast uncertainty from the full predictive distribution is valuable information to be considered by practitioners for decision-making and one step to enhance the use of ANNs for decision-making (Gagliardi et al., 2017; Kley-Holsteg & Ziel, 2020).

A major advantage of these models is however, that they are able to model non-linear relationships between the response variable and its predictors and are thus well capable to mimic the complex and stochastic nature of the demand itself (Menapace, Zanfei & Righetti, 2021). Therefore, in combination with the computational efficiency and ease of implementation, this study proposes a probabilistic deep neural network (DNN) for short-term water demand prediction focusing particularly on top days. The developed models are used in combination with the extraction of the feature importance, which thus not only produce accurate forecasts but also help to gain insights on the water demand as well as the influencing factors in context. Lastly, due to the data-driven approach, the presented models show a good transferability, allowing the implementation for multiple water utilities as it is demonstrated in the study.

17.2 Data

Developing ML models heavily relies on access to sufficient data of high quality. This study uses real data which is aggregated on district level and stems from two water-supplying companies in the central and western part of Germany (see Figure 17.1). The raw data is carefully pre-processed so that outliers are removed, missing data is imputed and clock change is adjusted. As a data-driven approach is chosen, we use other variables, that help describing water demand patterns as additional predictors to improve the forecast performance. These are finally described together with the proposed model in Section 18.3.1.

Multiple studies on water demand forecasting confirm that the past water demand is the most important predictor for the future water demand (Xenochristou et al., 2021; Zubaidi et al., 2018). Furthermore, water demand holds multiple seasonalities (see Figure 17.1). As previous analytics by Xenochristou and Kapelan (2020) prove, water demand is higher during the summer months, when water is used for outdoor activities (see Figure 17.1a), as well as weekends (see Figure 17.1b), when people tend to spend more time at home. In addition, water use follows a diurnal pattern during the day, with peak consumption during the morning (7–8 a.m.) and evening (6–8 p.m.) hours (see Figure 17.1c), when most people wake up or come back from work, respectively.

Reviewing water demand research approaches including weather variables, air temperature is most frequently named, to have a stronger effect on water use (Fiorillo et al., 2021; Manouseli, Kayaga & Kalawsky, 2019; de Souza et al., 2015). Weaker effects are found between water demand and the occurrence and the intensity of rainfall (Chang, Praskievicz & Parandvash, 2014; de Souza et al., 2015; Xenochristou et al., 2020).

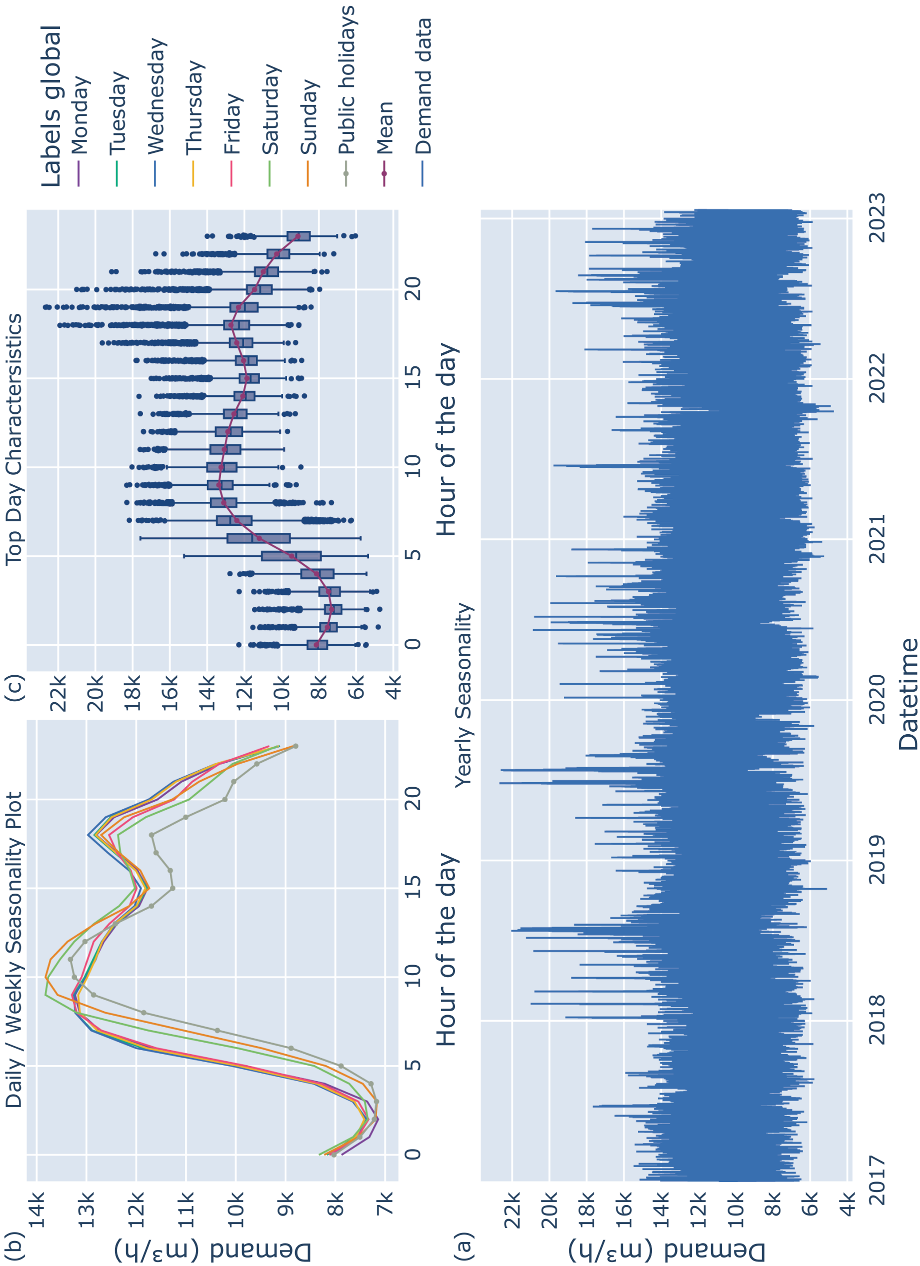


Figure 17.1: Plot of exemplified water demand data: (a) time series plot, (b) daily periodicity plot and (c) top day periodicity plot which present the days with highest demand. (Compiled by authors.)

As estimating the non-linear effect of weather on water demand could be of particular importance on top days, this study puts emphasise on analysing the effects through a temporally disaggregated analysis. To do so, measured weather data of multiple stations that lie within the supply areas provided by German Weather Service is used to compute the correlation to the water demand during top days.

These are the top days within the dataset where the demand peaks have been sampled using a combination of methods including expert knowledge, quantile analysis and residual diagnostics. To avoid missing a single top day, a joint of the resulting days stemming from the different sampling methods was selected for further analysis. Table 17.1 presents the segmented evaluation of the Spearman's rank correlation which is chosen to determine statistical significances of the relationships between weather data and water demand, segregated by different temporal characteristics such as time of day, day of week and season on top days as similarly done by Xenochristou (2019). Likewise, we choose Spearman's rank correlation to assess the degree of monotonic relationships, since it is better suited to identify non-linear relationships (Myers, Well & Lorch, 2010).

Table 17.1: Spearman's ρ correlation coefficients for water demand and weather variables (air temperature, sunshine duration, cumulative dry hours, relative humidity, soil temperature). (Compiled by authors.)

Temporal feature	Segmentation category	Air temperature	Sunshine duration	Cumulative dry hours	Relative humidity	Soil temperature
Time of day	Morning (6–12 a.m.)	0.403	0.376	0.346	−0.306	**0.409**
	Afternoon (12–6 p.m.)	0.608	0.279	0.393	−0.432	**0.640**
	Evening (6–12 p.m.)	**0.712**	0.620	0.156	−0.599	0.653
	Night (12–6 a.m.)	0.185	NaN	0.032	0.043	**0.311**
Day of week	Workdays	0.634	**0.680**	0.204	−0.579	0.509
	Weekends (plus special days)	0.684	**0.726**	0.141	−0.617	0.510
Season	Spring	0.574	**0.745**	0.220	−0.511	0.419
	Summer	0.676	**0.694**	0.198	−0.641	0.555
	Autumn	0.406	**0.661**	0.225	−0.394	0.289

The main findings from the table suggest that there is a moderate to strong positive correlation between water demand and air temperature for all temporal characteristics. The highest correlation is found during the evening of top days (6–12 p.m.) with a coefficient of 0.71, which suggests that water demand is sensitive to changes in air temperature. Through additional investigation, we could furthermore state that the effect of air temperature is especially noticeable for temperatures above 20°C. Similarly, there is a moderate to strong positive correlation between water demand and sunshine duration in evenings and all other segmentation categories for day of the week and season.

While most high correlation coefficients are attributed to summer water use, sunshine duration is the only weather variable that correlates better with consumption during the spring season with a coefficient of 0.75. On the other hand, there is a moderate negative correlation between water demand and relative humidity for all temporal characteristics. The highest correlation is found during the summer period with a coefficient of –0.64, which suggests that as the relative humidity decreases, the water demand increases. In general, lower correlation values correspond to morning and nightly water use while difference between workdays and weekends seems to be neglectable. Cumulative dry hours are hardly correlated to the water demand for all temporal characteristics. Overall, this chapter's results can help forecasting water demand by considering the impact of weather on consumption, specifically by creating advanced water demand prediction models which can model the non-linear relationships found here.

17.3 Methodology

17.3.1 Deep Neural Network Model

DNNs are a type of ML algorithm that are modelled after the structure and function of the human brain. They allow complex non-linear relationships between the response variable and its predictors (Hyndman & Athanasopoulos, 2021). A (feed-forward) DNN consists of at least two interconnected neurons that transmit and process information. The neurons are organised as layers as exemplified in Equation (1) (Marcjasz et al., 2022). Let $X \in \mathbb{R}^{D \times N}$ be the input matrix to the net with N denoting the number of features and D the number of observations, the neurons in one layer process the input matrix from the previous layer, and transmit the output matrix $H_i \in \mathbb{R}^{D \times h_i}$ to connected neurons in the subsequent layer, which again has adjustable weights denoted as $W_i \in \mathbb{R}^{h_{i-1} \times h_i}$ and bias denoted as $b_i \in \mathbb{R}^{D \times h_1}$ that determine the significance of the input. The downstream neurons combine the input from all their weighted upstream connections. The i th activation function, denoted as $a_i(\cdot)$ process the summed inputs, producing results that propagate to either the next hidden layer or the output layer. As different layers may use distinct functions, these functions are crucial for defining the output of neurons in a DNN by applying them to the weighted sum of inputs and bias.

Finally, with $h_i \in \mathbb{N}$ being the number of neurons in the i th hidden layer with $h_0 = N$ and thus $H_0 = X$ one layer of a DNN for $i \in \{1, \ldots, I\}$ can be written as (Marcjasz et al., 2022):

$$H_i = a_i(W_i H_{i-1} + b_i) \tag{1}$$

For this study, the models are built and estimated using the TensorFlow (Abadi et al., 2016) and Keras (Chollet, 2015) frameworks. The hyperparameter optimisation is performed with the help of Optuna package, which employs a Bayesian optimisation algorithm, specifically the Tree-structured Parzen Estimator (TPE), during the tuning process (Akiba et al., 2019). In the present study, we tuned the number of layers, number of units, layer activations and the optimiser in a study with 1000 trials. The final model structure is illustrated in Figure 17.2. The information travels from the inputs (predictors) which form the top layer to the outputs (forecasts) which form the bottom layer as described above. The chosen multilayer feed-forward network architecture includes three hidden layers containing hidden neurons which add non-linearity through their rectified linear unit (ReLU) activation functions to the network (Hyndman & Athanasopoulos, 2021).

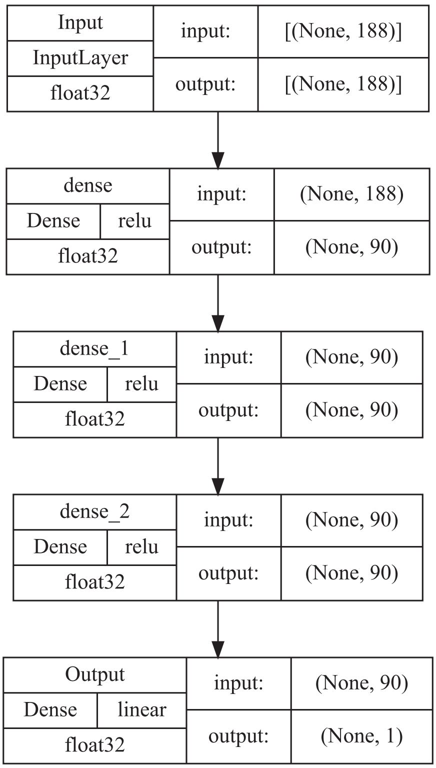


Figure 17.2: Final architecture of the Keras model instance including layer activations and dtypes. Adam is used as optimiser to compile the model. (Compiled by authors.)

As input features, we choose the first 168 lags of the water demand, temporal features such as sine and cosine waves for hour of the day, day of the week and week of the year, whether the day was a weekday or a weekend and whether it is a public holiday or a school holiday. Additionally, the models were equipped with values for different correlated meteorological features at different measuring stations that lie within the supply areas.

As the developed DNN models include lagged values of the water demand the network can more precisely called a DNN autoregression or $DNNAR_{(p)}$ model where p denotes the number of lagged inputs of the actual water demand.

When it comes to forecasting, the network is applied iteratively. For forecasting one step ahead, the model simply uses historical inputs. To run a multi-step ahead forecast, the one-step forecast is fed back as lagged demand to the input along with the historical data. This procedure is applied iteratively until the model has computed the complete forecast horizon, including the creating of prediction intervals as described in Section 18.3.2.1.

17.3.2 Forecasting Study Design

The forecasting performance of the proposed models is evaluated on both in-sample and out-of-sample data. The in-sample fit holds information on how well the training data is learnt by the model. The out-of-sample fit can be seen as an indicator of how well the trained model handles unseen data under operational conditions. While the desired model performs particularly well on top days while simultaneously also handles normal days satisfactory, we created a balanced dataset holding 40 forecasting steps consisting of 20 top days as well as 20 normal days to evaluate the performance on unseen data. It is worth noting that the forecasting time point starts every corresponding day at midnight. In this study, we always used at least two complete years covering 2017 and 2018 for training to cover at least one full year of data and to also train on a particular dry and hot year 2018 where many top days occurred.

17.3.2.1 Prediction Intervals from Bootstrapped Residuals

To determine the distribution of our forecast, we assume that the forecast errors are uncorrelated as in Hyndman and Athanasopoulos (2021). If $\hat{y}$ is the predicted value of the N-th sample, and y is the corresponding true value, then our forecast error is denoted as $\epsilon_t = y_t - \hat{y}_t$. The next observation of a time series can then be simulated as:

$$y_{T+1} = \hat{y}_{T+1|T} + \epsilon_{T+1} \tag{2}$$

where $\hat{y}_{T+1|T}$ is the one-step forecast and ϵ_{T+1} is the unknown future error. Based on the assumption that future errors will be similar to past errors, ϵ_{T+1} is replaced by a sampling from our in-sample residuals. An ensemble is then created by doing this repeatedly in a Monte Carlo simulation with a total of $M = 1000$ sample paths.

17.3.2.2 Benchmarks

To evaluate the proposed model's performance, several benchmark models were implemented, of which we present three in this study. It is to note that all benchmark models do not use external regressors like weather data. The first model consists of a relatively simple *Expert* model which fits the deseasonalised vector of the demand, including the fixed lags 1, 24 and 168 in a deseasonalised $AR(p)$ model. Furthermore, to provide another easily implementable and fast computable forecasting benchmark, one de-seasonalised $AR(p)$ model is implemented, which uses an integrated lag selection based on Bayesian-Information-Criterion (BIC).

Finally, a Lasso estimated high-dimensional time series model based on Kley-Holsteg and Ziel (2020) is applied which outperformed several other state-of-the-art methods significantly and which will be referred to as $ARXARCHX_{lasso}$ in the following. All the respective models are applied in a probabilistic way based on the sampling technique described in Section 18.3.2.1.

17.3.2.3 Forecast Evaluation Methods

An appropriate choice of evaluation methods with respect to the studied forecasting problem is important to quantify the regression performance with respect to the actual values (Ziel & Berk, 2019). All forecasting models produce probabilistic forecasts, which can be reduced to point forecasts by calculating the mean of $\hat{y}_t \in \mathbb{R}^H$ for all steps of the forecast horizon H, with t denoting the time of the first forecast horizon step. Thus, common error metrics like the mean absolute error (MAE) and the root mean squared error (RMSE) can be applied which provide good interpretability. If $\hat{y}$ is the predicted value of the N-th sample, and y is the corresponding true value, then the MAE estimated over N evaluated data points is defined as:

$$\text{MAE} = \frac{1}{NH} \sum_{t=1}^{N} \| y_t - \hat{y}_{t1} \| \tag{3}$$

And the RMSE is defined as:

$$\mathrm{RMSE} = \frac{1}{N}\sum_{t=1}^{N}\sqrt{\frac{1}{H}\,||y_t - \hat{y}_t||_2^2} \tag{4}$$

However, these evaluation methods discard the information on the predictive distribution. Here, the energy score (ES) (Gneiting & Raftery, 2007) is well suited because it takes into account the full distribution of forecast probabilities, rather than just the point forecast.

As a strictly proper scoring rule, the ES is estimated from M independently sampled trajectories $\tilde{y}_{tm}$ of the inflow over the forecast horizon with:

$$\mathrm{ES} = \frac{1}{M}\sum_{i=1}^{M} ||y_t - \tilde{y}_{t,m}||_2^{\beta} - \frac{1}{2M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} ||\tilde{y}_{t,i} - \tilde{y}_{t,j}||_2^{\beta} \tag{5}$$

where $\beta = 1$ according to Ziel and Berk (2019).

17.3.2.4 Significance Test

To enable statements on the statistical significance of one forecast reaching a higher accuracy than another forecasts, the Diebold Mariano (DM) test is used (Diebold & Mariano, 1995). The test holds the null hypothesis of no difference in the accuracy of two competing forecasts. Here a two-sided DM test is calculated with a significance level of 5% to test for significant differences from zero in both directions. For further details of the DM test, see Diebold and Mariano (1995).

17.4 Results and Discussion

This chapter presents the results of the developed $DNNAR_{(p)}$ model in comparison with the respective benchmarks for the in-sample and out-of-sample data and the respective forecast evaluation methods. Obtained scores for the out-of-sample period are presented in Table 17.2. Finally, the usefulness of complete probabilistic multi-step-ahead based on deep learning models is discussed and future research aims are outlined.

For all scores shown in Table 17.2, the implemented $ARXARCHX_{lasso}$ model still dominates the other models in terms of forecasting accuracy.

In this study, the $DNNAR_{(p)}$ models, both with and without weather data, emerged as the second-best performers, outperforming the conventional $AR(p)$ and the *Expert* models. Notably, the inclusion of weather data slightly enhanced the $DNNAR_{(p)}$ model's performance by 3.36% (ES) even though the correlation coefficients between certain weather conditions and water demand in Table 17.1 were partly significant. One

Table 17.2: Forecasting results and performance improvements (Imp.) in % relative to the *$ARXARCHX_{lasso}$* model for each forecasting model within the out-of-sample period. (Compiled by authors.)

Forecasting models	ES (m^3/h)	Imp. (%)	MAE (m^3/h)	Imp. (%)	RMSE (m^3/h)	Imp. (%)
Expert***	5663.90	–46.84	983.21	–40.95	1351.32	–30.91
$\boldsymbol{AR(p)}^*$	4282.38	–11.02	725.98	–4.07	1114.05	–7.93
$\boldsymbol{ARXARCHX_{lasso}}^*$	3857.24	0.00	697.58	0.00	1032.22	0.00
$\boldsymbol{DNNAR(p)}^*$	4064.13	–5.36	718.58	–3.01	1068.19	–3.48
$\boldsymbol{DNNAR(p)}^{**}$	3923.81	–1.73	699.51	–0.28	1034.12	–0.18

* Without weather data

** With weather data

explanation might be that even in the $DNNAR_{(p)}$ model equipped with weather variables, the most important features that improved the model performance were the previous states (discharge measurements in the past) and daily and hourly periodicity. A gradual increase of the water demand over time attributed to a rise in temperature would partially already be captured by those features. This also explains the overall high scores on top days whose dynamics remain difficult to capture for all proposed models. Overall, it is worth noting that checking the statistical significance of the higher accuracy of the $ARXARCHX_{lasso}$, the null hypothesis of the DM test could not be rejected when compared to the $DNNAR_{(p)}$ models.

While the linear models like the $ARXARCHX_{lasso}$ inherently adapt well, they struggle to capture non-linear relationships with exogenous variables. In contrast, the $DNNAR_{(p)}$ models allow to learn complex dependencies to explanatory variables (e.g., weather information) from the data but are less effective in forecasting non-stationary data.

Consequently, it is very likely, that a hybrid approach combining $ARXARCHX_{lasso}$ and $DNNAR_{(p)}$ model including weather data could potentially offer an even more robust predictive accuracy and improvements over the individual models as done for example by Berrisch, Narajewski & Ziel (2023) in the field of electricity demand prediction. This synergy might be beneficial even if the neural network is significantly worse but still delivers reasonable results as observed by De Vilmarest and Goude (2022).

Furthermore, while the ReLU activation function excels in introducing non-linearity and promoting sparse activations, it may not be optimal for modelling water demand in relation to weather features that exhibit both positive and negative correlations. A different activation function that captures negative activations, such as Leaky ReLU or Tanh, might provide a more nuanced representation of such contrasting relationships.

Finally, Figure 17.3 presents a visual overview of the predictive capability of the $DNNAR_{(p)}$ forecasting model. The figure displays two out-of-sample forecasts, one for a normal day and the other for a top day. Regarding the normal day forecast (on the left), we can observe that the model captured the morning hours quite well while slightly overestimating in the afternoon. Upon comparing the top day forecast, the model better predicts the morning peak (when people tend to wake up later) which can be attributed to the incorporation of weather data.

Overall, this research shows that DNNs have been a popular method for solving a variety of problems, including forecasting water demand. This is largely due to their computational efficiency and ease of implementation. The architecture of DNNs is relatively simple, yet it can be adjusted to handle complex problems and produce probabilistic outputs, making it a versatile tool for decision-making.

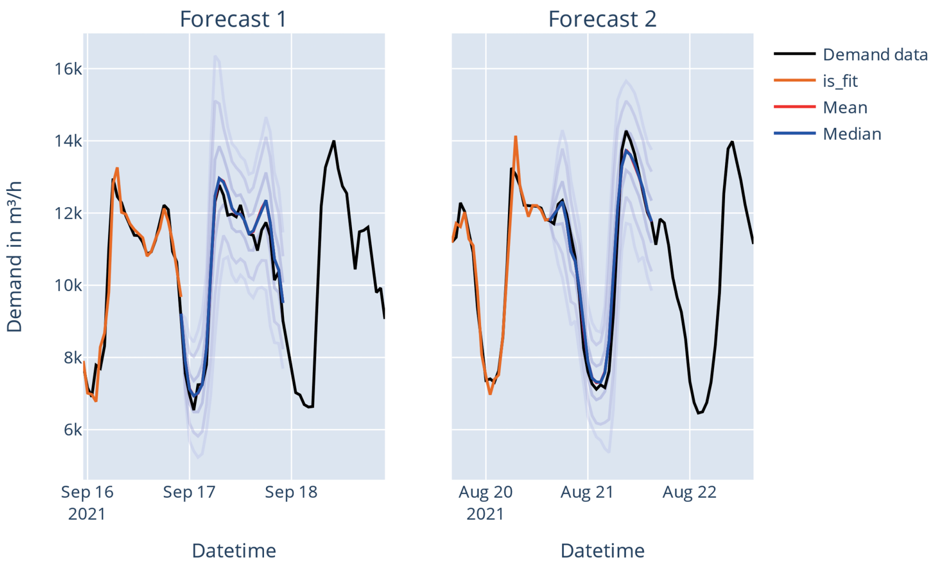


Figure 17.3: Plot of two complete probabilistic multi-step-ahead forecasts on Friday, 17 September 2021 (left) and Saturday, 21 August 2018 (right). (Compiled by authors.)

Additionally, the data-driven approach of these models allows them to generalise well to other water demand data. This is especially important for forecasting water demand, as water usage patterns can vary significantly between different supply regions and seasons. As more data becomes available for training, we can expect the quality of the model to improve.

One potential avenue for further improvement is to apply an autoregressive recurrent neural network (RNN). RNNs proved to be effective for time series forecasting in many disciplines as they can memorise past inputs, which allows them to capture patterns in the data that may not be immediately apparent. Furthermore, by applying an autoregressive RNN model, one clear advantage to this style of model is that it can be set up to produce output with a varying length. Furthermore, incorporating actual weather forecasts, rather than assuming perfect forecasts based on historical data, would also provide a more realistic assessment of the uncertainty added to the model in a production environment. Future work will focus on the incorporation of meteorological forecasts instead of historical values for weather data as well as the additional uncertainty connected to it.

Bibliography

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., . . . Zheng, X. (2016). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems* (arXiv:1603.04467). arXiv. https://doi.org/10.48550/arXiv.1603.04467.

Adamowski, J. F. (2008). Peak daily water demand forecast modeling using artificial neural networks. *Journal of Water Resources Planning and Management*, *134*(2), 119–128. https://doi.org/10.1061/(ASCE)0733-9496(2008)134:2(119).

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631. https://doi.org/10.1145/3292500.3330701.

Berrisch, J., Narajewski, M., & Ziel, F. (2023). High-resolution peak demand estimation using generalized additive models and deep neural networks. *Energy and AI*, *13*, 100236. https://doi.org/10.1016/j.egyai.2023.100236.

Chang, H., Praskievicz, S., & Parandvash, H. (2014). *Sensitivity of Urban Water Consumption to Weather and Climate Variability at Multiple Temporal Scales: The Case of Portland, Oregon*. *International Journal of Geospatial and Environmental Research*, *1*(1), 7.

Chollet, F., & others (2015). *Keras*. GitHub. Retrieved January 15, 2023, from https://github.com/fchollet/keras.

de Souza, A., Aristone, F., Sabbah, I., Santos, D. A. da S., Lima, A. P. de S., & Lima, G. (2015). Climatic variations and consumption of urban water. *Atmospheric and Climate Sciences*, *5*(3), Article 3. https://doi.org/10.4236/acs.2015.53022.

de Souza Groppo, G., Costa, M. A., & Libânio, M. (2019). Predicting water demand: A review of the methods employed and future possibilities. *Water Supply*, *19*(8), 2179–2198. https://doi.org/10.2166/ws.2019.122.

De Vilmarest, J., & Goude, Y. (2022). State-space models for online post-Covid electricity load forecasting competition. *IEEE Open Access Journal of Power and Energy*, *9*, 192–201. https://doi.org/10.1109/OAJPE.2022.3141883.

Diebold, F., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.

Donkor, E. A., Mazzuchi, T. A., Soyer, R., & Alan Roberson, J. (2014). Urban water demand forecasting: Review of methods and models. *Journal of Water Resources Planning and Management*, *140*(2), 146–159. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000314.

Fiorillo, D., Kapelan, Z., Xenochristou, M., De Paola, F., & Giugni, M. (2021). Assessing the impact of climate change on future water demand using weather data. *Water Resources Management*, *35*(5), 1449–1462. https://doi.org/10.1007/s11269-021-02789-4.

Gagliardi, F., Alvisi, S., Kapelan, Z., & Franchini, M. (2017). A probabilistic short-term water demand forecasting model based on the markov chain. *Water*, *9*(7), Article 7. https://doi.org/10.3390/w9070507

Ghalehkhondabi, I., Ardjmand, E., Young, W. A., & Weckman, G. R. (2017). Water demand forecasting: Review of soft computing methods. *Environmental Monitoring and Assessment*, *189*(7), 313. https://doi.org/10.1007/s10661-017-6030-3.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. https://doi.org/10.1198/016214506000001437.

Guo, G., Liu, S., Wu, Y., Li, J., Zhou, R., & Zhu, X. (2018). Short-term water demand forecast based on deep learning method. *Journal of Water Resources Planning and Management*, *144*(12), 04018076. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000992.

Hyndman, R., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. Retrieved October 19, 2022, from https://research.monash.edu/en/publications/forecasting-principles-and-practice-3.

Kley-Holsteg, J., & Ziel, F. (2020). Probabilistic multi-step-ahead short-term water demand forecasting with lasso. *Journal of Water Resources Planning and Management*, *146*(10), 04020077. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001268.

Kühnert, C., Gonuguntla, N. M., Krieg, H., Nowak, D., & Thomas, J. A. (2021). Application of LSTM networks for water demand prediction in optimal pump control. *Water*, *13*(5), Article 5. https://doi.org/10.3390/w13050644.

Manouseli, D., Kayaga, S. M., & Kalawsky, R. (2019). Evaluating the effectiveness of residential water efficiency initiatives in England: Influencing factors and policy implications. *Water Resources Management*, *33*(7), 2219–2238. https://doi.org/10.1007/s11269-018-2176-1.

Marcjasz, G., Narajewski, M., Weron, R., & Ziel, F. (2022). *Distributional neural networks for electricity price forecasting* (arXiv:2207.02832). arXiv. http://arxiv.org/abs/2207.02832

Menapace, A., Zanfei, A., & Righetti, M. (2021). Tuning ANN hyperparameters for forecasting drinking water demand. *Applied Sciences*, *11*(9), Article 9. https://doi.org/10.3390/app11094290.

Myers, J.L., Well, A.D., & Lorch Jr, R.F. (2010). *Research Design and Statistical Analysis* (3rd ed.). Routledge. https://doi.org/10.4324/9780203726631.

Tiwari, M. K., & Adamowski, J. F. (2015). Medium-term urban water demand forecasting with limited data using an ensemble wavelet–bootstrap machine-learning approach. *Journal of Water Resources Planning and Management*, *141*(2), 04014053. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000454.

Xenochristou, M. (2019). *Water demand forecasting using machine learning on weather and smart metering data*. Retrieved January 15, 2023, from https://ore.exeter.ac.uk/repository/handle/10871/39792.

Xenochristou, M., Hutton, C., Hofman, J., & Kapelan, Z. (2020). Water demand forecasting accuracy and influencing factors at different spatial scales using a gradient boosting machine. *Water Resources Research*, *56*(8), e2019WR026304. https://doi.org/10.1029/2019WR026304.

Xenochristou, M., Hutton, C., Hofman, J., & Kapelan, Z. (2021). Short-term forecasting of household water demand in the UK using an interpretable machine learning approach. *Journal of Water Resources Planning and Management*, *147*(4), 04021004. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001325.

Xenochristou, M., & Kapelan, Z. (2020). An ensemble stacked model with bias correction for improved water demand forecasting. *Urban Water Journal*, *17*(3), 212–223. https://doi.org/10.1080/1573062X.2020.1758164.

Ziel, F., & Berk, K. (2019). *Multivariate Forecasting Evaluation: On Sensitive and Strictly Proper Scoring Rules* (arXiv:1910.07325). arXiv. https://doi.org/10.48550/arXiv.1910.07325.

Zubaidi, S. L., Gharghan, S. K., Dooley, J., Alkhaddar, R. M., & Abdellatif, M. (2018). Short-term urban water demand prediction considering weather factors. *Water Resources Management*, *32*(14), 4527–4542. https://doi.org/10.1007/s11269-018-2061-y.

List of Contributors

David Alaminos
Universitat de Barcelona
E-mail: alaminos@ub.edu

Andrea Altepost
RWTH Aachen University
E-mail: a.altepost@wirksam.nrw

Wolfgang Arens-Fischer
University of Applied Sciences Osnabrück
E-mail: w.arens-fischer@hs-osnabrueck.de

Alexander Bönner
Ruhr West University of Applied Sciences
E-mail: alexander.boenner@hs-ruhrwest.de

Marcus Bravidor
University of Freiburg
Email: marcus.bravidor@accounting.uni-freiburg.de

Ed Dandalt
Wilfried Laurier University
E-mail: e.dandalt@wlu.ca

Timm Eichenberg
Hochschule Weserbergland
E-mail: eichenberg@hsw-hameln.de

Arne Eimuth
Ruhr West University of Applied Sciences
E-mail: arne.eimuth@hs-ruhrwest.de

Ingo Elsen
FH Aachen University of Applied Sciences
E-mail: elsen@fh-aachen.de

Markus Feld
University of Applied Sciences Osnabrück
E-mail: m.feld@hs-osnabrueck.de

Alexander Ferrein
FH Aachen University of Applied Sciences
E-mail: ferrein@fh-aachen.de

Adjan Hansen-Ampah
RWTH Aachen University
E-mail: a.hansen-ampah@wirksam.nrw

Markus Harlacher
Institute of Applied Industrial Engineering and Ergonomics, Düsseldorf
E-mail: m.harlacher@wirksam.nrw

Janis Hummel
HWZ University of Applied Sciences in Business Administration Zurich
E-mail: janis.hummel@fh-hwz.ch

Oliver Isoz
V-ZUG AG, Zug
E-mail: oliver.isoz@gmx.ch

Gregor Johnen
University of Duisburg-Essen
E-mail: gregor.johnen@uni-due.de

Jens Kley-Holsteg
University of Duisburg-Essen

Medina Klicic
Ruhr West University of Applied Sciences
E-mail: medina.klicic@hs-ruhrwest.de

Katharina I. Köstner
Universitat de Barcelona
E-mail: 31koestk7@alumnes.ub.edu

Jacqueline Krüger
University of Osnabrück
E-mail: jacque.krueger@gmail.com

Nicolai Krüger
IU International University of Applied Science Information Systems
E-mail: nicolai.krueger@iu.org

Isabel Lausberg
Ruhr West University of Applied Sciences
E-mail: isabel.lausberg@hs-ruhrwest.de

https://doi.org/10.1515/9783110790320-018

Amanda Lemette
Pontificia Universidade Catolica do Rio de Janeiro
E-mail: amanda.lemette@puc-rio.br

Bàrbara Llacay
Universitat de Barcelona
E-mail: bllacay@ub.edu

Khuliso Mapila
University of Johannesburg
E-mail: 200908385@student.uj.ac.za; khully22@gmail.com

Robert J. Martin
University of Duisburg-Essen
E-mail: robert.martin@uni-due.de

Richard von Maydell
ETH Zürich
E-mail: rbaron@ethz.ch

Robert Menger
Menger Engineering GmbH, Leipzig
E-mail: robert@menger.group

Christoph Menzel
Federal Ministry for Economic Affairs and Climate Action, Berlin
E-mail: christoph.menzel@bmwk.bund.de

Wolfgang Merx
RWTH Aachen University
E-mail: w.merx@wirksam.nrw

Tankiso Moloi
University of Johannesburg
E-mail: smoloi@uj.ac.za

Simon Moser
HWZ University of Applied Sciences in Business Administration Zurich
E-mail: simon.moser@fh-hwz.ch

Sina Niehues
Institute of Applied Industrial Engineering and Ergonomics, Düsseldorf
E-mail: s.niehues@wirksam.nrw

André Niemann
University of Duisburg-Essen

Anne Stockem Novo
Ruhr West University of Applied Sciences
E-mail: anne.stockem-novo@hs-ruhrwest.de

Karla Ohler-Martins
Ruhr West University of Applied Sciences
E-mail: karla.ohler-martins@hs-ruhrwest.de

Thomas Pitz
Rhine-Waal University of Applied Sciences
E-mail: thomas.pitz@hochschule-rhein-waal.de

Nils Pudill
Ruhr West University of Applied Sciences
E-mail: nils.pudill@hs-ruhrwest.de

Zunera Rana
Radboud University
E-mail: zuneraahmad.rana@ru.nl

Thomas Rautenstrauch
Eastern Switzerland
University of Applied Sciences
E-mail: thomas.rautenstrauch@ost.ch

Jessica Roemer
Rhine-Waal University of Applied Sciences
E-mail: jessica.roemer@hochschule-rhein-waal.de

Michael Römmich
Ruhr West University of Applied Sciences
E-mail: michael.roemmich@hs-ruhrwest.de

Simone Roth
Ruhr West University of Applied Sciences
E-mail: simone.roth@hs-ruhrwest.de

Britta Rüschoff
FOM Hochschule für Oekonomie und Management
E-mail: britta.rueschoff@fom.de

Stefan Schiffer
FH Aachen University of Applied Sciences
E-mail: s.schiffer@fh-aachen.de

Marcel Schumacher
University of Applied Sciences Osnabrück
E-mail: marcel.schumacher@hs-osnabrueck.de

Fatemeh Nasim Shahinfar
Institute of Applied Industrial Engineering and Ergonomics, Düsseldorf
E-mail: f.shahinfar@wirksam.nrw

Joern Sickmann
Rhine-Waal University of Applied Sciences
E-mail: joern.sickmann@hochschule-rhein-waal.de

Irina Simon
Ruhr-West University of Applied Sciences
E-mail: irina.simon@hs-ruhrwest.de

Agnis Stibe
RMIT University Vietnam
E-mail: agnis@agnisstibe.com

Merih Türkoglu
Ruhr West University of Applied Sciences
E-mail: merih.tuerkoglu@hs-ruhrwest.de

Michael Vogelsang
Ruhr West University of Applied Sciences
E-mail: michael.vogelsang@hs-ruhrwest.de

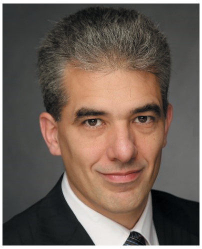
Jan Vogt
sipgate GmbH, Düsseldorf
E-mail: contact@j-vogt.com

Malte Weiß
Ruhr West University of Applied Sciences
E-mail: malte.weiss@hs-ruhrwest.de

Florian Ziel
University of Duisburg-Essen

About the Editors

Isabel Lausberg is Professor of Business and Management Accounting at the Ruhr West University of Applied Sciences (HRW) in Mülheim an der Ruhr, Germany. As co-director of the research focus “Artificial Intelligence from an Economic Perspective”, she focuses on digitalization and AI in her research and teaching. She is a certified data scientist, specialized in data analytics, and has been recognized as a fellow for innovations in digital higher education. Prior to joining HRW, Lausberg was a professor at the Technical University of Applied Sciences in Bochum (THGA) and gained practical business experience in e-commerce and strategic planning in a leading German retail company. She holds a degree in Business from the University of Cologne and a PhD from the University of Duisburg-Essen. Foto: HRW / Köhring



Michael Vogelsang is Professor of Economics at the Ruhr West University of Applied Sciences (HRW) in Mülheim an der Ruhr, Germany. His academic focus encompasses economics, digitalization, and quantitative methods. In 2020, he, together with Isabel Lausberg, established a pioneering research group dedicated to “Artificial Intelligence from an Economic Perspective”. Now, as intelligence becomes scalable, technological innovations are merging with economic drivers. The interdisciplinary perspective on AI, ranging from computer science to philosophy, is a key part of his fascination with the subject. Prior to his current role, he was Chief Economist at the German Association of Small and Medium-sized Enterprises in Berlin. Vogelsang's academic career includes a PhD from the University of Wuppertal, where he specialized in the study of digitalization in open economies.

https://doi.org/10.1515/9783110790320-019

List of Figures

https://doi.org/10.1515/9783110790320-020

List of Tables

https://doi.org/10.1515/9783110790320-021