Influencing Factors on the Adoption of AI: Insights From Social, Organizational, Individual and Methodological Perspectives

TECHNISCHE UNIVERSITÄT DARMSTADT

Department of Law and Economics
at the Technical University of Darmstadt

Dissertation

by

Maren Felicitas Mehler

submitted in fulfilment of the requirements for the degree of
Doctor rerum politicarum (Dr. rer. pol.)

First Assessor:	Prof. Dr. Peter Buxmann
Second Assessor:	Prof. Dr. Alexander Benlian
Darmstadt	2025

Mehler, Maren Felicitas

Influencing Factors on the Adoption of AI: Insights From Social, Organizational, Individual and Methodological Perspectives

Darmstadt, Technische Universität Darmstadt,

Year thesis published in TUprints 2025

Date of the viva voce: 27/01/2025

Declaration of Authorship

I hereby declare that the submitted thesis is my own work. All quotes, whether word by word or in my own words, have been marked as such.

The thesis has not been published anywhere else nor presented to any other examination board.

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig angefertigt habe. Sämtliche aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Arbeit wurde bisher weder einer anderen Prüfungsbehörde vorgelegt noch veröffentlicht.

Maren Felicitas Mehler

Darmstadt, October 9, 2024

Abstract

Artificial Intelligence (AI) applications are becoming increasingly important, simplifying daily life and supporting organizations across various applications. Despite the numerous positive attributes and the potential of AI, studies indicate that adoption rates, particularly within organizations, are still not as high as expected. To increase the utilization of AI, factors such as the user's culture, industry-specific elements, willingness to pay, and psychological factors play a critical role. By leveraging factors that promote AI adoption and addressing barriers, it is possible to enhance the integration of AI technologies. This dissertation examines the factors influencing AI adoption from four perspectives: (1) social, (2) organizational, (3) individual, and (4) methodological.

From the *social perspective*, one study included in this dissertation investigates the influence of culture on the adoption of AI as an emerging technology among others. A structured literature review (SLR) was conducted, focusing on Information Systems (IS) papers from the Basket of Eight that measure the effect of culture. The knowledge extracted from these papers was then condensed, with existing research categorized by research areas, data collection methods, and their assessments. The resulting concept matrix serves as a valuable summary and foundation for future research. A recommendation from the SLR is that future research should measure culture individually rather than making assumptions based on the country a person lives in. Additionally, the study explicitly provides a research agenda highlighting existing gaps in the literature. For instance, the cultural impact on the adoption of newer forms of AI, such as Generative AI (GenAI), should be measured. Thus, this study demonstrates the significant influence of culture as a social factor on AI adoption.

For the *organizational perspective*, two studies were conducted. The first study utilizes a case study approach, consisting of seven interviews within the financial services and manufacturing industries. From these interviews, drivers and barriers of AI adoption were identified and categorized using the Technology-Organization-Environment (TOE) framework. The factors were also classified by industry and compared between the two industries. This leads to the identification of soft factors that are industry-specific and hard factors that are more general. For example, a soft barrier specific to the financial industry is the presence of legacy systems, while a general driver is the potential for cost reduction. The identified factors are particularly useful for organizations within these industries, but the more general factors can be applied to other organizations as well. The second study within the organizational perspective examined the willingness to pay (WTP) for machine learning-based software testing tools in organizations using a conjoint analysis. Initially, attributes important for these tools and their target audience were identified through a structured literature review and a Delphi study. Attributes such as accuracy, ease of use, and integration were found to be crucial for the adoption. The conjoint analysis, conducted with 119 software testers in Germany, revealed that they are willing to pay up to €120 more per license per month for an increase in accuracy from 90% to 99%. This study highlights WTP as an adoption factor when introducing AI and identifies the essential attributes AI systems must possess to be successfully adopted. Thus, the organizational perspective uncovers various influencing factors examined within this dissertation.

Also, two studies were conducted from the *individual perspective*. Both utilized online experiments to investigate the psychological factors influencing AI adoption. The first study in this perspective examined the impact of ChatGPT assistance on the performance and perceived meaningfulness among programmers. The study involved 161 experienced coders who completed coding and debugging tasks with and without ChatGPT assistance. The results showed a significant increase in performance but a decrease in perceived meaningfulness due to the reduced difficulty of the tasks. As the results are depending on the tasks, the adoption of AI should be carefully considered as lower meaningfulness can result in less motivation for work. Another study in this perspective, involving 174 participants from Germany, demonstrated that the IKEA effect occurs when using Generative AI. This effect arises when individuals value an output more if they invested more effort in its creation. This suggests that instead of solely aiming for automation, which is often seen as the goal of AI, fostering collaboration could enhance AI adoption. These studies collectively demonstrate the significant influence of individual psychological factors on AI adoption.

Finally, the *methodological perspective* emphasizes the importance of thorough and sound methodology. Using a Design Science Research approach, this study initially examines how previous research has conducted and reported cross-sectional surveys. Based on this analysis, focus groups were utilized to gain deeper insights. This results in eleven guidelines which serve as a foundation to allow particularly inexperienced researchers to conduct their (AI adoption) research in a reproducible and reliable manner.

Overall, these studies show that various factors influence the adoption of AI, which must be considered from different perspectives. Only by examining these different angles AI can be successfully implemented. Thus, AI does not only simplify daily life but also supports organizations in improving or streamlining processes. Moreover, the role of society is also crucial, as well as a solid methodology to make research generalizable.

Abstract (German Version)

Künstliche Intelligenz (KI) gewinnt zunehmend an Bedeutung und erleichtert nicht nur das tägliche Leben, sondern unterstützt auch Unternehmen in verschiedenen Anwendungsbereichen. Trotz der zahlreichen positiven Eigenschaften und des Potenzials von KI zeigen Studien, dass die Adoptionsraten, insbesondere in Unternehmen, noch hinter den Erwartungen zurückbleiben. Um die Nutzung von KI zu steigern, spielen Faktoren wie die Kultur der Anwender, industriespezifische Eigenschaften, die Zahlungsbereitschaft und psychologische Aspekte eine entscheidende Rolle. Durch die gezielte Nutzung von Treibern und die Beseitigung von Hindernissen kann die Adoption von KI-Technologien vorangetrieben werden. Diese Dissertation untersucht die Einflussfaktoren der KI-Adoption daher aus vier Perspektiven: (1) gesellschaftlich, (2) unternehmerisch, (3) individuell und (4) methodisch.

Aus der *gesellschaftlichen Perspektive* beleuchtet eine der in dieser Dissertation enthaltenen Studien den Einfluss von Kultur auf die Adoption von KI als eine von vielen neuen Technologien. Dazu wurde eine strukturierte Literaturübersicht durchgeführt, die sich auf die Wirtschaftsinformatik-Forschung beschränkt und Artikel aus dem „Basket of Eight" analysierte, die den Effekt von Kultur messen. Das aus diesen Artikeln gewonnene Wissen wurde anschließend kondensiert, wobei die vorhandene Forschung nach Forschungsbereichen, Datenerhebungsmethoden und deren Auswertung kategorisiert wurde. Die daraus entstandene Konzeptmatrix dient als Zusammenfassung und Grundlage für zukünftige Forschung. Eine Empfehlung aus der Literaturrecherche ist, dass zukünftige Forschung Kultur individuell messen sollte, anstatt Annahmen basierend auf dem Land, in dem eine Person lebt, zu treffen. Darüber hinaus liefert die Studie eine explizite Forschungsagenda, die bestehende Lücken in der Literatur aufzeigt. So sollte beispielsweise der kulturelle Einfluss auf die Adoption neuerer Formen der KI, wie generative KI, untersucht werden. Diese Studie zeigt somit den bedeutenden Einfluss der Kultur als gesellschaftlicher Faktor auf die KI-Adoption.

Für die *unternehmerische Perspektive* wurden zwei Studien durchgeführt. Die erste Studie nutzte einen Case Study Ansatz, bei dem sieben Interviews in der Finanzdienstleistungs- und Fertigungsindustrie geführt wurden. Aus diesen Interviews wurden Treiber und Barrieren der KI-Adoption identifiziert und mithilfe des Technology-Organization-Environment (TOE) Frameworks kategorisiert. Die Faktoren wurden außerdem nach Industrien geordnet und innerhalb der Industrien verglichen, was zur Identifikation von weichen Faktoren führte, die industriespezifisch sind, und harten Faktoren, die allgemeiner gelten. Ein spezifisches Beispiel für eine weiche Barriere in der Finanzindustrie ist die alte IT Infrastruktur (Legacy Systeme), während ein allgemeiner harter Treiber die potenzielle Kostenreduktion ist. Die identifizierten Faktoren sind besonders nützlich für Unternehmen innerhalb dieser Industrien, die allgemeineren Faktoren können jedoch auch auf andere Organisationen angewendet werden. Die zweite Studie innerhalb der unternehmerischen Perspektive untersuchte die Zahlungsbereitschaft für ML-basierte Software Testing Tools in Unternehmen mithilfe einer Conjoint-Analyse. Zunächst wurden durch eine strukturierte Literaturrecherche und eine Delphi-Studie Attribute identifiziert, die für diese Tools und ihre Zielgruppe wichtig sind. Attribute wie Genauigkeit, Benutzerfreundlichkeit und Integration erwiesen sich als entscheidend für die Adoption. Die Conjoint-Analyse, die mit 119 Softwaretestern in Deutschland durchgeführt wurde, zeigte, dass die Teilnehmer bereit sind, bis zu 120 € mehr pro Lizenz pro Monat für eine Erhöhung der Genauigkeit von 90 % auf 99 % zu zahlen. Diese Studie hebt die Zahlungsbereitschaft als

Adoptionsfaktor bei der Einführung von KI hervor und identifiziert die wesentlichen Eigenschaften, die KI-Systeme besitzen müssen, um erfolgreich adoptiert zu werden. Die unternehmerische Perspektive deckt somit verschiedene Einflussfaktoren auf, die in dieser Dissertation untersucht wurden.

Auch aus der *individuellen Perspektive* wurden zwei Studien durchgeführt. Beide nutzten Online-Experimente, um psychologische Faktoren zu untersuchen, die die KI-Adoption beeinflussen. Die erste Studie in dieser Perspektive untersuchte den Einfluss der Unterstützung durch ChatGPT auf die Leistung und die wahrgenommene Sinnhaftigkeit bei Programmierern. Die Studie umfasste 161 erfahrene Programmierer, die jeweils eine Codierungs- und eine Debugging-Aufgabe mit und ohne ChatGPT-Unterstützung durchführten. Die Ergebnisse zeigten eine signifikante Leistungssteigerung, jedoch auch eine Verringerung der wahrgenommenen Sinnhaftigkeit aufgrund der geringeren Aufgabenschwierigkeit. Da die Ergebnisse abhängig von den Aufgaben sind, sollte die Einführung von KI sorgfältig abgewogen werden, da eine geringere Sinnhaftigkeit zu einer geringeren Arbeitsmotivation führen kann. Eine weitere Studie in dieser Perspektive, an der 174 Teilnehmer aus Deutschland teilnahmen, zeigte, dass auch der IKEA-Effekt bei der Nutzung von generativer KI auftritt. Dieser Effekt tritt auf, wenn Personen einen Output mehr wertschätzen, wenn sie mehr Aufwand in dessen Erstellung investiert haben. Dies deutet darauf hin, dass anstatt allein auf Automatisierung, die oft als Ziel von KI angesehen wird, die Förderung der Zusammenarbeit die KI-Adoption erhöht werden sollte. Diese Studien zeigen insgesamt den erheblichen Einfluss individueller psychologischer Faktoren auf die KI-Adoption.

Schließlich betont die *methodische Perspektive* die Bedeutung einer gründlichen und soliden Methodik. In einem Design Science Research-Ansatz untersucht diese Studie zunächst, wie bisherige Forschung Cross-Sectional Surveys durchgeführt und berichtet hat. Basierend auf dieser Analyse wurden Fokusgruppen durchgeführt, um tiefere Einblicke zu gewinnen. Dies führte zu elf Richtlinien, die als Grundlage dienen, um insbesondere unerfahrenen Forschern zu ermöglichen, ihre Forschung zur (KI-Adoption) reproduzierbar und zuverlässig durchzuführen.

Insgesamt zeigen diese Studien, dass verschiedene Faktoren die Adoption von KI beeinflussen, die aus unterschiedlichen Perspektiven betrachtet werden müssen. Nur durch die Betrachtung dieser verschiedenen Blickwinkel kann KI erfolgreich implementiert werden, was nicht nur das tägliche Leben erleichtert, sondern auch Unternehmen unterstützt und deren Prozesse verbessert oder vereinfacht. Auch die Rolle der Gesellschaft ist entscheidend, und um Forschung generalisierbar zu machen, ist eine solide Methodik unerlässlich.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ABM	Agent-based modeling
AI	Artificial intelligence
ANOVA	Analysis of variance
AVE	average variance extracted
B	Barriers
BMBF	German Federal Ministry of Education and Research
CE	Circular economy
CIOs	Chief Information Officer
CMB	Common method bias
CR	composite reliability
CX	Customer experience
D	Drivers
DOI	Diffusion of Innovation
DSR	Design Science Research
DSS	Decision Support Systems
E	Environment
Ex	Expert
ECIS	European Conference on Information Systems
EU	European Union
FS	Financial services
G	General
G #	Guideline number
GDP	Gross domestic product
GDPR	General Data Protection Regulation
GenAI	Generative Artificial Intelligence
H	Hypothesis
HICSS	Hawaii International Conference on System Sciences
HR	Human Resource
HTMT	Heterotrait-monotrait ratio
I	Interviewee
ICIS	International Conference on Information Systems
IDV	Individualism vs. Collectivism
IP	Inexperienced Participant

IS	Information Systems
IT	Information Technology
IVR	Indulgence vs. (Self-)Restraint
kNN	k-nearest neighbors
LLM	Large Language Model
LTO	Long Term Orientation
M (italic)	Mean
M	Manufacturing
MAS	Masculinity vs. Femininity
MC	Mass customization
ML	Machine Learning
NLP	Natural Language Processing
O	Organization
P	Proficient participants
PACIS	Pacific Asia Conference on Information Systems
PDI	Power Distance Index
Ph.D.	Doctor of Philosophy
PLS	Partial Least Square
PTKA	Project Management Agency Karlsruhe
ROI	return on investment
RQ	Research Question
SD	Standard deviation
SLR	Structured Literature Review
SME	Small and Medium Enterprise
T	Technology
TAM	Technology Acceptance Model
TOE	Technology-Organization-Environment
TPB	Theory of planned behavior
TRA	Theory of reasoned action
UAI	Uncertainty Avoidance Index
UK	United Kingdom
US	United States
UTAUT	Unified Theory of Acceptance and Use of Technology
VHB	Verband der Hochschullehrerinnen und Hochschullehrer für Betriebswirtschaft e.V.
WTP	Willingness to pay
XAI	Explainable Artificial Intelligence

1 Introduction

1.1 Overarching Motivation

Artificial Intelligence (AI) has gained increasing significance in recent years, both in our daily life and in various professional contexts. From voice assistants and translations that assist us with everyday tasks (e.g., DeepL, 2023; McElheran et al., 2024; Subhash et al., 2020), to personalized recommendations in online shops (e.g., Krishnan & Mariappan, 2024), and advanced analytical tools in medicine (e.g., Ellenrieder et al., 2023; Pumplun et al., 2021)—the possibilities of AI are almost limitless (Dwivedi et al., 2023; Russell & Norvig, 2021). This rapid development is primarily driven by continuous advancements in key technologies. As early as the 2000s, machine learning laid the foundation for today's AI revolution by enabling machines to learn from data and continuously improve their performance. A few years later, the breakthrough in deep learning, a specialized form of machine learning that utilizes neural networks allowed for deeper and more complex data analysis. This led to significant progress in areas such as image and speech recognition (Berente et al., 2021; Erik Brynjolfsson & Mitchell, 2017; Mitchell, 1997; Russell & Norvig, 2021).

The next major step in the development of AI technologies occurred in November 2022 with the introduction of ChatGPT and the widespread public awareness of Generative AI (GenAI) (Dwivedi et al., 2023; OpenAI, 2022). GenAI systems and especially Large Language Models (LLMs), which are capable of generating new content, have fundamentally changed the way we interact with machines. These systems are not only used for automating routine tasks but also in creative processes, such as the creation of texts, images, or music (Dwivedi et al., 2023; Feuerriegel et al., 2024; Teubner et al., 2023). Overall, these developments have led to AI being utilized in a wide range of industries today, whether in healthcare to support diagnosis, in finance for fraud detection, or in coding for software development (Bao et al., 2022; Dwivedi et al., 2023; Ellenrieder et al., 2023). The diversity and breadth of AI applications underscore its enormous potential to enhance efficiency and innovation across nearly all areas of economic and social life.

However, before technologies like AI can fully realize their potential, they must first be integrated into the organizational processes and accepted by users (Radhakrishnan & Chattopadhyay, 2020). This process of adoption, which refers to the uptake and utilization of new technologies, is a complex and multifaceted phenomenon that occurs at the societal, organizational, and individual levels (Kumar et al., 2024; McElheran et al., 2024; Radhakrishnan & Chattopadhyay, 2020; Rogers, 2003; Uzumcu & Acilmis, 2024). At the societal level, adoption can be influenced by frameworks such as regulatory adjustments and cultural changes. For example, regulations that facilitate the use of AI technologies or ensure their safety can accelerate the spread of AI (Alzebda & Matar, 2024). At the organizational level, readiness for adoption is often reflected in investments in new technologies. Companies must not only allocate financial resources but also create the necessary infrastructure and organizational structures to effectively integrate the technology. This often requires changes to internal processes and adjustments to corporate strategy (Peters et al., 2020; Pumplun et al., 2019; Sturm et al., 2021). Finally, at the individual level, the adoption of end-users ultimately determines whether and how new technologies are actually utilized. The introduction of AI often requires users to acquire new skills and adapt their workflows. This individual adoption is influenced by various factors, such as trust in the technology, perceived usefulness, and ease of use (Bedué & Fritzsche, 2022; Radhakrishnan & Chattopadhyay, 2020; Sturm & Peters, 2020).

Studies shows that the interest in AI has significantly increased in recent years, as evidenced by rising investments and a growing number of AI applications across various industries. From less than 1% in 2016 the market interest has grown to 17% (Goldman Sachs, 2023) with investments expected up to 826 billion US Dollars in 2030 (Statista, 2024). Also, AI can improve customer relationships or increase productivity and thus can counteract labor shortages (Haan, 2023; IBM, 2022a). Nevertheless, the adoption rate of AI remains below expectations in many areas. For example in 2023 a survey in the EU only found an adoption rate of 8% among organizations (Eurostat, 2023). This is partly due to specific challenges associated with AI technologies, such as the lack of explainability in many AI models, often referred to as the "Black Box" problem (Berente et al., 2021). This opacity can undermine user trust, thereby hindering the acceptance and integration of the technology. Another obstacle is the high demand for high-quality data necessary for the successful deployment of AI (Polisetty et al., 2024; Pumplun et al., 2019). In data-scarce environments, this can significantly complicate the implementation and utilization of AI (Jiang et al., 2022; Korteling et al., 2021; Polisetty et al., 2024). In addition to these known challenges, there are numerous other factors that influence AI adoption and are thus critical to the success and full realization of this technology's potential. These factors include technical infrastructure, organizational readiness for change, ethical and legal considerations, as well as the individual attitudes and skills of users (Pumplun et al., 2019).

1.2 Research Questions

This dissertation aims to systematically investigate and analyze various influencing factors to understand how they affect the adoption of AI technologies at different levels and from different perspectives. Through this analysis, the goal is to develop a deeper understanding of the conditions under which AI can be successfully integrated into organizations and societies, and to identify the strategies necessary to overcome existing barriers. The adoption of AI is a multifaceted process influenced by various factors that can either drive or hinder its integration. Moreover, AI adoption occurs at multiple levels, making it necessary to approach the topic from a comprehensive perspective. In this dissertation, four distinct perspectives are identified, which together form the overarching research objective: **How is AI adoption influenced from (1) the social perspective, (2) the organizational perspective, (3) the individual perspective, and (4) the methodological perspective**. Each perspective can be addressed through a specific research question (RQ).

When considering the question of AI adoption, it is logical to begin with the broadest perspective—the social one. Using AI has a high impact on society, but the public also influences the adoption of AI (Floridi & Cowls, 2019) The adoption of emerging technologies, such as AI, is influenced by numerous factors, including national culture, which plays a critical role (e.g., D. J. Kim et al., 2016; Kumar et al., 2024; Shore & Venkatachalam, 1996). Emerging technologies like AI offer a competitive advantage for companies (Wulf et al., 2017) and can simplify both life and work (Berente et al., 2021; Dwivedi et al., 2023; Russell & Norvig, 2021; A. Stahl, 2021). Numerous studies have demonstrated that national culture is a decisive factor in the successful introduction of emerging technologies, making it crucial to investigate its impact (Ives & Jarvenpaa, 1991; Myers & Tan, 2002; Shore & Venkatachalam, 1996). Thus, the first research question:

RQ1: How do social factors influence the adoption of artificial intelligence at the national level?

AI adoption can also be examined at a more granular level, either from an individual perspective or from an organizational standpoint. Particularly in business, organizations can gain numerous benefits from the utilization of AI, such as enhancing efficiency, improving decision-making, for example through supporting software development (Ali et al., 2023; Bhalerao et al., 2022; Shrestha et al., 2019; Venkatesh, 2022). Moreover, new forms of AI, such as GenAI, are increasingly capable of supporting and taking over additional tasks like writing texts or generating images (Dwivedi et al., 2023). However, research has long recognized that the adoption of AI itself is often the primary obstacle to the effective utilization of these technologies (Brynjolfsson & McAfee, 2017). Several studies have already identified factors that promote or hinder AI adoption within organizations (e.g., Kruse et al., 2019; Pumplun et al., 2019). These studies have also noted that industry-specific analysis can reveal specific adoption factors that organizations in particular industries should consider (Cubric, 2020; Kar et al., 2021; Pumplun et al., 2019; Zöll et al., 2022). On the other hand, the success of AI adoption within organizations also depends on how AI systems are implemented. Given the capabilities of pre-trained AI services like ChatGPT, the willingness of organizations to invest in these technologies is a critical factor that must be taken into account. This willingness is often dependent on the features and functionalities that the AI system offers (Peters et al., 2020). This leads to the second research question:

RQ2: What organizational factors promote or hinder the successful adoption of artificial intelligence in companies?

Whether in a corporate setting as an employee or as a private individual, psychological characteristics also play a role in AI adoption. Research in human-AI collaboration has shown that the cooperation between humans and machines is crucial for the successful adoption of technology (e.g., Abel-Karim et al., 2020; Boyacı et al., 2023; Fügener et al., 2022; Jussupow et al., 2021). However, GenAI is increasingly taking over tasks that were once the exclusive domain of humans, such as creativity or empathy (Bryant, 2023). While this expansion of capabilities can free up human capacity for other tasks, it can also lead to perceptions of AI as a threat, thereby hindering its acceptance and adoption. Additionally, cognitive biases may influence decisions to adopt AI. For instance, research on algorithm aversion demonstrates that people are often reluctant to follow AI-generated recommendations or decisions (Jussupow et al., 2020). Conversely, biases can also have positive effects and can be leveraged to facilitate AI adoption. This leads to the third research question:

RQ3: Which factors determine the willingness of individuals to adopt artificial intelligence in their professional and private lives?

In addition to AI adoption itself, a robust methodological approach is crucial for any research endeavor. Only when research is conducted rigorously and systematically the results can be reliable and generalizable. This is particularly true for quantitative research, which forms the basis of several papers in this dissertation. Quantitative methods, especially surveys, allow for the observation of phenomena across a broad sample and the statistical validation of relationships, provided they are conducted properly (Dinev et al., 2013; Melnyk et al., 2012; Recker, 2021). However, these methods must be carefully developed and executed to ensure their added value. Therefore, the final research question aims to establish a foundation for AI adoption research using cross-sectional surveys. This is addressed by the fourth research question:

RQ4: How can methodological approaches be utilized to measure the adoption of artificial intelligence?

In summary, these four research questions are interconnected and collectively address the multifaceted nature of AI adoption (see Figure 1-1). They provide a comprehensive framework for understanding how AI can be successfully integrated into various contexts and for identifying the strategies necessary to overcome the challenges associated with its adoption.

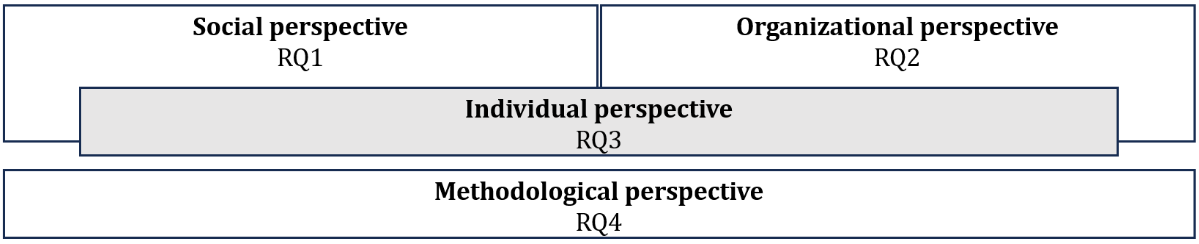

Figure 1-1 – Overview of Relation Between RQs

1.3 Structure of This Dissertation

To address the research questions outlined above, this dissertation comprises six peer-reviewed conference papers, which are summarized in Table 1-1.

Research Question	Paper	Citation
RQ1: Social perspective	Paper A	Mehler, M., Turan Akdag, M., & Zöll, A. (2023). Exploring the Effect of National Culture on Emerging Technologies: A Glimpse into the Future. PACIS 2023 Proceedings. **Methodology**: Structured Literature Review **VHB-Ranking**[1]: C
RQ2: Organizational perspective	Paper B	Hoffmann, M., & Mehler, M. (2023). An Industry-Specific Investigation on Artificial Intelligence Adoption: The Cases of Financial Services and Manufacturing. PACIS 2023 Proceedings. **Methodology**: Case Study **VHB-Ranking**: C
	Paper C	Mehler, M. F., & Vetter, O. A. (2023). How Much Are Machine Assistants Worth? Willingness to Pay for Machine Learning-Based Software Testing. ECIS 2023 Proceedings. **Methodology**: Conjoint Analysis **VHB-Ranking**: A
RQ3: Individual perspective	Paper D	Mehler, M., & Krautter, K. (2024). Productivity vs. Purpose: Generative AI Enhances Task Performance but Reduces Meaningfulness in Programming. ECIS 2024 Proceedings. **Methodology**: Online Experiment **VHB-Ranking**: A

[1] This dissertation is based on the latest VHB ranking (see VHB, 2024) for assessing the quality of peer-reviewed papers.

Research Question	Paper	Citation
	Paper E	Mehler, M., Ellenrieder, S., & Buxmann, P. (2024). The Influence of Effort on the Perceived Value of Generative AI: A Study of the IKEA Effect. ECIS 2024 Proceedings. **Methodology**: Online Experiment **VHB-Ranking**: A
RQ4: Methodological perspective	Paper F	Mehler, M., Ellenrieder, S., Turan Akdag, M., Wagner, A., & Benbasat, I. (2023). A Framework for Developing Cross-Sectional Surveys. ICIS 2023 Proceedings. **Methodology**: Design Science Research **VHB-Ranking**: A

Table 1-1 – List of Publications Included in This Dissertation

Paper A focuses on the first research question by examining the influence of national culture on the adoption of emerging technologies like AI from a social perspective. Through a structured literature review following the methodology of vom Brocke et al. (2009), this paper investigates how the effect of national culture on the usage and adoption of emerging technologies has been measured in information systems (IS) research. A total of 28 papers from the Basket of Eight journals were analyzed based on their application scenarios, the cultural theories employed, the methods of data collection, data analysis techniques, and other factors such as publication year. Based on this analysis, a research agenda was developed to identify existing gaps and suggest how future research should measure culture, emphasizing the need to assess culture at the individual level to capture different cultural influences and personal preferences. The findings demonstrate that culture is a significant factor in the adoption and usage of emerging technologies like AI, and it should be considered in future research, particularly with respect to Generative AI.

Paper B addresses the second research question by exploring the organizational perspective of AI adoption. This paper examines the factors influencing AI adoption in the financial services and manufacturing industries. Through seven interviews conducted within a case study methodology, the paper analyzes which factors promote and hinder AI adoption. Using the Technology-Organization-Environment (TOE) framework, the study identifies 20 drivers of AI adoption, such as cost reduction, competitive pressure, and curiosity, as well as 30 barriers, including legacy IT, data availability, and a lack of AI understanding, across both industries. These factors are further categorized into hard (generalizable) and soft (industry-specific) AI adoption factors. This paper provides valuable insights into the organizational factors that companies must consider when adopting AI and offers a transparent guide to the insights and knowledge of the AI industry.

Paper C also relates to the second research question and focuses on an essential aspect of AI adoption: the willingness to pay. The premise is that organizations must be willing to purchase AI technologies at a price they perceive as fair in order to adopt and implement these technologies effectively. The paper explores this issue in the context of ML-based software testing. First, a Delphi study was conducted to identify the attributes and levels that influence users' willingness to pay and, consequently, their adoption decisions. Following this, a choice-based conjoint analysis with 119 participants in Germany was carried out, explicitly targeting individuals in companies to gain an organizational perspective on adoption. The findings highlight that, in addition to the price of an AI product, accuracy—a typical measure of correctness in the ML context—has a significant impact on the choice of AI service. Integration of the AI tool and ease of use also play crucial roles. Thus, three key factors influencing AI adoption in organizations are emphasized.

Paper D focuses on the third research question, which examines the individual perspective of AI adoption. This paper investigates how the adoption of Generative AI influences both task performance and the perceived meaningfulness of tasks. The study takes a software engineering perspective, conducting an online experiment with 161 experienced programmers to assess the impact of ChatGPT assistance on programming and debugging tasks. The results indicate that ChatGPT facilitates tasks, leading to higher performance, which could positively influence AI adoption. However, the simplification of tasks also resulted in reduced perceived meaningfulness among individuals, potentially complicating or even preventing adoption. The type of task also influenced the successful use of Generative AI, highlighting the importance of considering individual psychological factors in AI adoption.

Paper E also addresses the third research question by examining the personal perspective of AI adoption through the lens of the IKEA effect. Using an online experiment, the study investigates whether the IKEA effect—where greater effort invested in a task leads to higher appreciation and perceived value of the results—applies to Generative AI. This effect could potentially enhance AI adoption if AI tools are used correctly. In the experiment, 174 participants were asked to complete one task collaboratively with Generative AI and another task solely by the AI. The results suggest that the effect is task-dependent, with the IKEA effect being particularly evident in tasks related to image generation. These findings indicate that collaboration with AI is a critical factor in AI adoption, and while automation may be beneficial in some contexts, human-AI interaction holds greater promise. This further emphasizes the significant role of individual psychology in AI adoption.

Finally, Paper F adopts the fundamental methodological perspective. In IS research, a solid methodological foundation is crucial for ensuring that research findings are transparent, replicable, and valid. Cross-sectional surveys, which are frequently used in AI adoption research (e.g., Abdalla et al., 2024; Mohr & Kühl, 2021; C. Wang et al., 2023), are a key methodology in this context. This paper adopts a Design Science Research approach, beginning with a structured literature review to identify best practices. Building on these insights, initial guidelines for developing cross-sectional surveys were created and then validated and refined through focus groups. This paper provides valuable guidance for future research on AI adoption, ensuring that cross-sectional surveys are conducted rigorously and scientifically, thereby enhancing the contributions of this dissertation.

This dissertation is structured as follows (see Figure 1-2). After the introduction, Chapter 2 provides an overview of the necessary foundations for this dissertation. Chapters 3 to 8 present the papers described above. The dissertation concludes with a discussion of the theoretical and practical contributions.

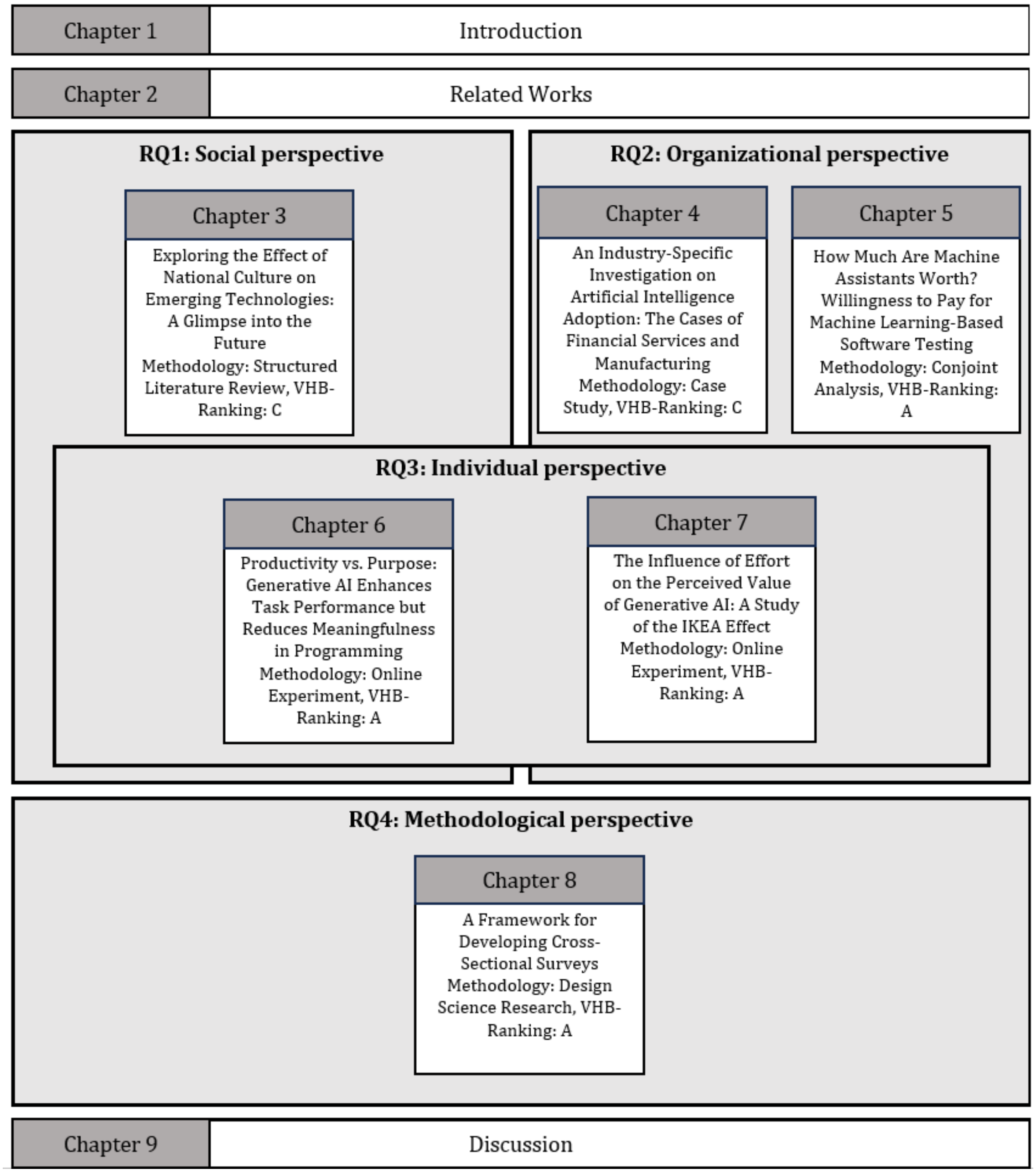

Figure 1-2 – Outline of This Dissertation

In addition to the papers presented in this dissertation, I have also contributed to other peer-reviewed conference papers, which are not part of this dissertation:

- Ellenrieder, S., Mehler, M., & Turan Akdag, M. (2023). Design for Acceptance and Intuitive Interaction: Teaming Autonomous Aerial Systems with Non-experts. PACIS 2023 Proceedings. 55. Methodology: Design Science Research, VHB-Ranking: C.
- Ellenrieder, S., Ellenrieder, N., Hendriks, P., & Mehler, M. (2024). Pilots and Pixels: A Comparative Analysis of Machine Learning Error Effects on Aviation Decision Making. ECIS 2024 Proceedings. Methodology: Online Experiment, VHB-Ranking A. Claudio Ciborra Award First Runner Up.

- Gräf, M., Mehler, M., & Ellenrieder, S. (2024). AI Strategy in Action: A Case Study on Make-or-Buy for AI-based Services. PACIS 2024 Proceedings. Methodology: Case Study, VHB-Ranking: C.
- Gräf, M., Mehler, M., & Ellenrieder, S. (2024). Crisis Management in the Metaverse: Designing Virtual Worlds for Real-Worlds Resilience. ICIS 2024 Proceedings. Methodology: Design Science Research, VHB-Ranking: A.
- Hendriks, P., Sturm, T., Mehler, M., & Buxmann, P. (2024). The Impact of Artificial Intelligence on the Evolution of Culture. ICIS 2024 Proceedings. Methodology: Simulation, VHB-Ranking: A.
- Jacquemin, P. H., Turan Akdag, M., Mehler, M., Wahl, N., & Buxmann, P. (2023). Are Organizations Ready for Metaverse? Identifying Influencing Factors for Initiating Metaverse in the Organizational Context. ICIS 2023 Proceedings. Methodology: Interviews, VHB-Ranking: A.
- Jacquemin, P., Gräf, M., Mehler, M., Walenta, P., Hendriks, P., & Buxmann, P. (2025). Exploring Mental Workload Across Different Levels of Immersion: A Metaverse Perspective. HICSS 2025 Proceedings, VHB-Ranking: B.
- Unzicker, D., Mehler, M., Kammholz, L., Sturm, T., Ellenrieder, S., & Buxmann, P. (2024). All Eyes on the Reviewer: Understanding the Impact of GenAI on Mental Workload and Performance in Code Reviews. ICIS 2024 Proceedings. Methodology: Online Experiment, VBH-Ranking: A.
- Vetter, O. A., Mehler, M. F., & Buxmann, P. (2023). As Much Art as Science-Examining the Realization of Business Models Driven by Machine Learning Through a Dynamic Capabilities Perspective. ECIS 2023 Research Papers. Methodology: Interviews, VHB-Ranking: A.

2 Related Work

This chapter outlines the theoretical background of this dissertation and the included papers. In particular, it focuses on the fundamentals of AI, machine learning, and generative AI. In addition, the topic of AI adoption is considered, as well as influencing factors on the adoption of AI and the human-AI collaboration that results from that. Finally, the methodological foundations relevant to the papers are described.

2.1 Artificial Intelligence, Machine Learning, and Generative AI

Artificial Intelligence (AI) is currently one of the emerging technologies, as it is reaching a certain level of maturity with new developments such as ChatGPT, yet it continues to be permanently developed (Gartner, 2022; B. C. Stahl & Eke, 2024; Wheatley & Wilemon, 1999). Although AI is currently under continuous advancement, the term was coined as early as the 1960s, with initial approaches suggesting that AI enables machines to perform tasks that previously required human intelligence (Jiang et al., 2022). Thus, the aim of AI is to imitate human intelligence (Korteling et al., 2021; P. Wang, 2019). The concept of intelligence is crucial in defining AI. Intelligence is defined as the “ability to autonomously and efficiently achieve complex goals” (Korteling et al., 2021, p. 3).

Between the initial idea of artificial intelligence and today, there has been a constant fluctuation in the hype surrounding AI with the development of new approaches, increased computational power, and the availability of data and coding libraries (e.g., Brynjolfsson & McAfee, A., 2017; Jiang et al., 2022). In recent years, AI has become available as a ready-to-use product, making it easier for organizations to adopt (Russel & Norwig, 2021). AI can be defined as “intelligent agents that receive percepts from the environment and take actions that affect that environment” (Russell & Norvig, 2021, p. 7). Additional definitions exist, such as AI being the frontier of emerging technologies focusing on human intelligence for complex decision-making (Berente et al., 2021). Overall, it can be summarized that AI supports automation and information processing with constantly new approaches (Benbya et al., 2021; Berente et al., 2021). AI is particularly used to generate strategic advantages, encompassing all areas of an organization where efficiency can be increased to enhance performance and reduce costs (Bhalerao et al., 2022).

The application areas of AI are diverse. AI can be used in all industries, altering their core processes and changing the value proposition or even business models (Brynjolfsson & McAfee, 2017; Sjödin et al., 2021). Additionally, AI can facilitate work (Dwivedi et al., 2023) or reduce inequality (Noy & Zhang, 2023). For example, AI is used in medicine (e.g., in radiology (Rao et al., 2023)), cybersecurity (Prasad et al., 2023), marketing or social media (Bhalerao et al., 2022; Chuma & De Oliveira, 2023), customer service (Bhalerao et al., 2022; Chuma & De Oliveira, 2023), logistics (Bhalerao et al., 2022), human resources, and administrative tasks (Korzynski et al., 2023). AI also finds applications in software development, such as in software testing (Shafiq et al., 2021) or writing software code (E. Chen et al., 2023; Liu et al., 2017; Surameery & Shakor, 2023). Furthermore, AI can be used in systems analysis and design, suggesting multiple IT architectural designs and swiftly adapting configurations (Russo, 2024). Overall, AI can lead to higher productivity if its potentials are utilized (Ebert, 2024; Russo, 2024). These diverse applications can be utilized by both large organizations and SMEs (Bhalerao et al., 2022).

AI possesses several unique characteristics compared to traditional software. These characteristics are well summarized by Berente et al. (2021) as autonomy, learning, and inscrutability. Autonomy implies that AI operates independently without human intervention. Using data, the AI algorithm learns autonomously (Berente et al., 2021). Additionally, AI can improve itself through the nature of its training (Y. S. Lee et al., 2022). However, this can lead to a certain level of inscrutability (Berente et al., 2021). This is mainly due to the statistical probabilities generated by self-learning (Brynjolfsson & McAfee, A., 2017). This results in the so-called black-box property of machine learning, making it difficult for humans to comprehend the results and impossible to prove that a system works for all cases (Bauer et al., 2021; Brynjolfsson & McAfee, A., 2017). Errors are also challenging to detect and can already be present in the data, thus amplifying biases in the output (Brynjolfsson & McAfee, A., 2017; Lund, 2023b).

Moreover, AI can be categorized into narrow and general AI. General AI is the ultimate goal as it can process comprehensive tasks. However, as of today, only narrow AI (also known as weak, specialized, and limited AI) functions effectively, meaning AI systems often outperform humans in particular, predefined tasks (Jiang et al., 2022; Korteling et al., 2021). Furthermore, AI encompasses several specific subcategories, such as machine learning (ML), see Figure 2-1. ML, as a subcategory of AI, involves learning from patterns found in data (Russel & Norwig, 2021). This means humans do not need to provide all explanations as in conventional software development; instead, the ML algorithm learns to perform a task from examples (Brynjolfsson & McAfee, A., 2017). The ML algorithm creates a model based on the data (Russel & Norwig, 2021). From this pattern recognition, relationships can subsequently be derived for recommendations or predictions (Brynjolfsson & Mitchell, 2017; Mitchell, 1997; Russell & Norvig, 2021). ML uses various algorithms to discover these relationships. The so-called deep learning (also known as deep neural networks) is a subcategory that uses neural networks (Brynjolfsson & McAfee, A., 2017; Shinde & Shah, 2018). Neural networks consist of many layers and parameters modeled after the human brain (Korteling et al., 2021; Shinde & Shah, 2018). This structure allows it to process unstructured data such as text, images, or audio data in large quantities, as long as they are encoded accordingly (Brynjolfsson & McAfee, A., 2017; Jiang et al., 2022). Thus, the larger the dataset, the more examples the algorithm has to learn from (Brynjolfsson & McAfee, A., 2017). With sufficient data, the algorithm can generalize and autonomously provide outputs for new data (Jiang et al., 2022). Deep learning algorithms often output either a value in the context of a prediction or a classification into a predefined category (Jiang et al., 2022). However, a disadvantage of these deep learning algorithms is the difficulty in understanding how decisions are made, resulting in a lack of transparency and the need for explainability (Korteling et al., 2021).

ML algorithms can further be divided into supervised, unsupervised, and reinforcement learning. This classification describes both the data required and the potential output. In supervised learning, each row in the dataset needs a so-called label, which represents an assignment or result. Precise predictions and estimations can be provided as outputs. However, labels are not always available due to high costs from domain experts, the labor of labeling datasets, or the impossibility of labeling. In such cases, unsupervised learning algorithms like clustering can be applied to independently divide datasets into groups (Brynjolfsson & McAfee, A., 2017; Jiang et al., 2022; Russel & Norvig, 2021). Although humans are more akin to unsupervised learners, it is challenging to replicate this (Brynjolfsson & McAfee, A., 2017). The third type, reinforcement learning, differs significantly. Here, an agent can perform various actions within an environment. The agent is rewarded for performing actions well getting closer to the defined goal. This is useful for applications where a goal is set, but the path to achieve it is not predefined (Brynjolfsson & McAfee, A., 2017; Jiang et al., 2022).

Additionally, in recent years, Generative AI (GenAI) has developed. Unlike traditional machine learning, this form of AI generates new, meaningful data based on identified relationships instead of merely analyzing them (Dwivedi et al., 2023; Feuerriegel et al., 2024; Teubner et al., 2023). This can involve text, images, or even code, music, or videos (e.g., ChatGPT, GitHub Copilot, Dall-E, Suno). The rise of GenAI has mainly been attributed to the large language model (LLM) ChatGPT since November 2022. ChatGPT and other LLMs like Google's Bard allow the generation of new texts using a prompt. GPT stands for Generative Pre-trained Transformer and utilizes transformer technology, which means neural networks that create predictions (Korzynski et al., 2023; Sauvola et al., 2024). The innovation lies mainly in the dialogue format, allowing follow-up questions while maintaining the same context (Chuma & De Oliveira, 2023). Other tools like Stable Diffusion or Dall-E 2 function similarly, except the output is an image. With these images and texts, a machine can now perform tasks that were previously assumed only humans could do, as they can now handle not only repetitive but also creative tasks (Bankins & Formosa, 2020; Sauvola et al., 2024).

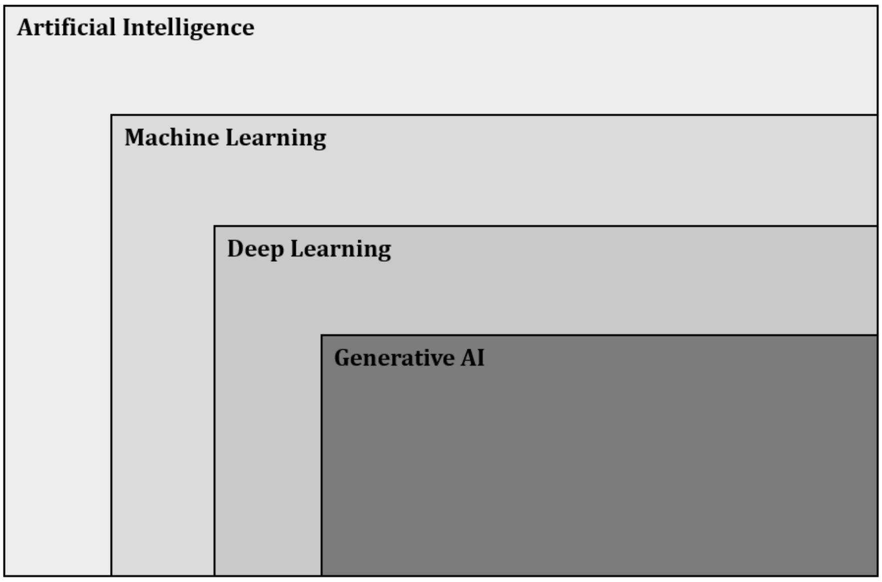

Figure 2-1 – Overview of the Relationship Between Artificial Intelligence, Machine Learning, Deep Learning, and Generative AI (Following Banh & Strobel, 2023)

AI can be utilized privately and professionally in almost all areas (Brynjolfsson & McAfee, A., 2017; Russel & Norvig, 2021). Particularly with the rise of GenAI, numerous new applications have emerged, as AI can now perform also creative tasks (Korzynski et al., 2023; Sauvola et al., 2024). However, the challenge remains that AI needs to be implemented effectively to realize its full potential.

2.2 AI Adoption

AI adoption falls under the research domain of technology diffusion and is examined within an organizational context concerning increased productivity (McElheran et al., 2024). Adoption refers to the implementation of a new technology within an organization as well as its utilization and acceptance (Radhakrishnan & Chattopadhyay, 2020).

Numerous models, frameworks, and theories exist to depict or measure technology adoption, which can also be applied to AI. These frameworks can assess adoption at the organizational or individual levels (Radhakrishnan & Chattopadhyay, 2020). Notable frameworks include the, the Diffusion of Innovations theory (DOI), the Technology Acceptance Model (TAM), the Technology-Organization-Environment (TOE) framework, and the Unified Theory of Acceptance and Use of Technology (UTAUT) (Ajzen, 1991; Davis, 1989; DePietro et al., 1990; Fishbein & Ajzen, 1975; Radhakrishnan & Chattopadhyay, 2020; Rogers, 2003; Venkatesh et al., 2003).

2.2.1 Diffusion of Innovations Theory

The Diffusion of Innovation (DOI) theory (see Figure 2-2) is recognized as a complex framework, primarily focusing on gathering information about the adaptation, innovation, and mitigating uncertainty (Agarwal et al., 1998). According to Rogers (2003), innovation is essential for generating alternative solutions, addressing both individual and institutional needs (Uzumcu & Acilmis, 2024). Rogers' (2003) definition of innovation is an object, idea, or practice that is perceived as new.

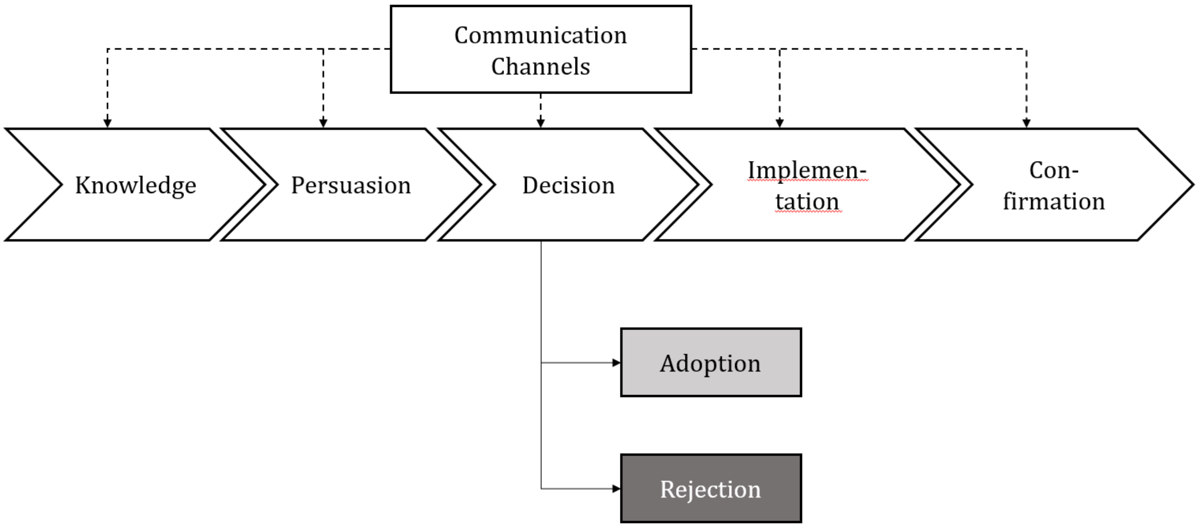

Figure 2-2 – Diffusion of Innovations Theory (Following Rogers, 2003)

DOI posits that individuals do not adopt innovations simultaneously but rather over varying periods (Lampo, 2022). Rogers (2003) identifies five distinct categories of adopters (see Figure 2-3): innovators, early adopters, early majority, late majority, and laggards. Innovators are characterized as creative, entrepreneurial individuals willing to take risks and embrace new ideas. Early adopters serve as role models for innovation adoption and are respected within their communities. The early majority are more cautious and deliberate in their adoption decisions. Conversely, the late majority are skeptical about innovations and typically adopt them only after a significant portion of society has done so. Finally, laggards are conservative and resistant to change, often exhibiting prejudice against new ideas (Uzumcu & Acilmis, 2024).

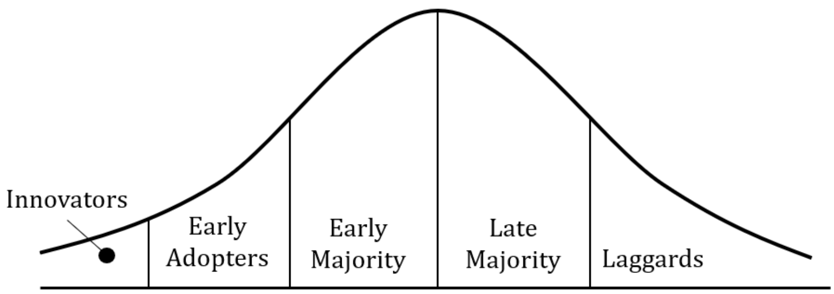

Figure 2-3 – Categories of Adopters According to DOI Theory (Following Rogers, 2003)

The DOI theory has been applied in various contexts to measure the adoption of new technologies. For instance, Abdalla et al. (2024) utilized the theory to assess the adoption of ChatGPT among students. Similarly, the theory has been employed to study the diffusion of smartphones in India (Wani & Ali, 2015). These examples illustrate the theory's applicability in understanding how different groups within a population adopt innovations over time.

2.2.2 Theory of Reasoned Action and Theory of Planned Behavior

The Theory of Reasoned Action (TRA), formulated by Fishbein and Ajzen in 1975, posits that the intention to perform a behavior is often automatic, consistent and reasonable from a set of beliefs about that behavior (Fishbein & Ajzen, 1975, 2011; Lampo, 2022). This theoretical framework has been extensively applied across various domains beyond technology to predict and explain virtually any human behavior. The core premise of TRA is that the performance of a behavior is the result of the intention to engage in that behavior. This intention is influenced by two primary predictors: attitude and subjective norm concerning the behavior in question (Fishbein & Ajzen, 1975, 2011; Lampo, 2022; Momani & Jamous, 2017).

Attitude refers to the evaluative positive or negative feelings associated with performing a behavior. For instance, if an individual believes that engaging in a particular behavior will lead to favorable outcomes, their attitude towards that behavior will be positive. On the other hand, the subjective norm pertains to the perceived expectations of significant others, such as family members, friends, or colleagues, regarding whether an individual should or should not perform a specific behavior. Thus, if significant others endorse or support the behavior, the individual is more likely to form a positive intention toward performing it (Fishbein & Ajzen, 1975).

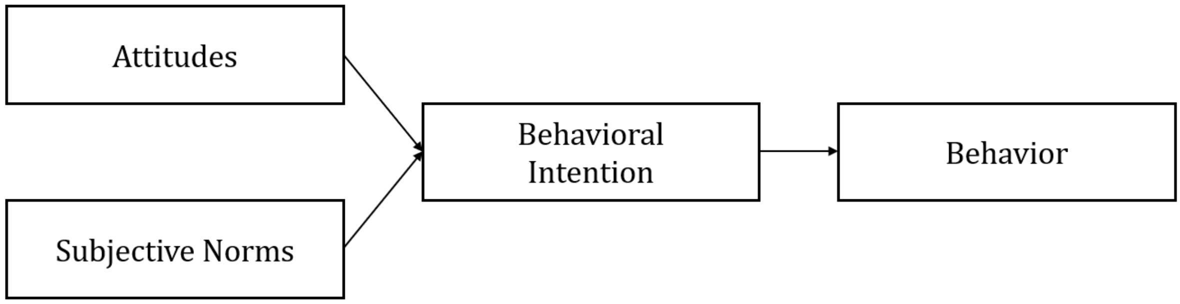

Figure 2-4 – Theory of Reasoned Action (Following Fishbein & Ajzen, 1975; Lampo, 2022)

The TRA is recognized as one of the most fundamental theories of human behavior and has inspired extensive theoretical research. Its significance lies in its systematic approach to understanding the motivational influences on behavior, emphasizing the role of intention as a central determinant (Lampo, 2022; Momani & Jamous, 2017). TRA was used, for example, to analyze the adoption of robo-advisors in fintech (Roh et al., 2023) or the use of the voice assistant on iPhones in the United Arab Emirates (Farhi et al., 2023).

To address scenarios where behavior performance requires specific resources (e.g., time, money, skills, or cooperation from others), the TRA was extended into the Theory of Planned Behavior (TPB). The TPB incorporates an additional construct: perceived behavioral control. This construct represents the perceived ease or difficulty of performing a specific behavior, reflecting an individual's assessment of the availability of resources and opportunities. Like TRA, the TPB suggests that behavioral intention is the strongest predictor of actual behavior. By acknowledging the role of perceived control, the TPB provides a more comprehensive understanding of the factors influencing behavior, particularly in contexts where resource constraints are significant (Ajzen, 1991).

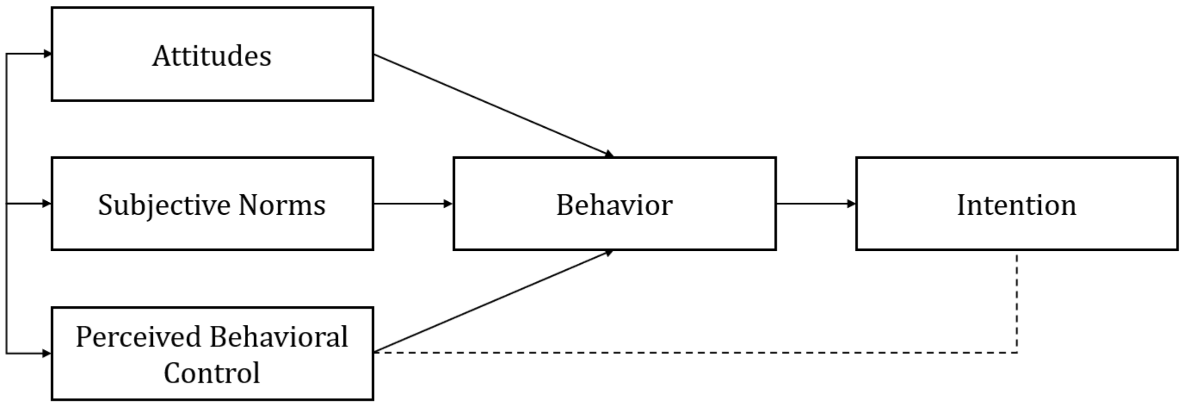

Figure 2-5 – Theory of Planned Behavior (Following Ajzen, 1991)

TPD is also applied broadly. For example, Mohr & Kühl (2021) investigate the acceptance of AI among German farmers, as do Sohn & Kwon (2020) on the acceptance of AI-based products, both utilizing the TPD.

2.2.3 Technology Acceptance Model

To explain the determinants of computer acceptance, the behavior of computer users and user populations, the Technology Acceptance Model (TAM) was initially proposed by Davis in 1991 adapted from the TRA (Theory of Reasoned Action) (Davis, 1989; Lampo, 2022). TAM uses the two technology-specific measures perceived usefulness and perceived ease of use, compared to the determinants of behavior in TRA. Perceived usefulness refers to the degree to which an individual believes using a particular system would enhance job performance. Conversely, perceived ease of use reflects the degree to which an individual believes that using a specific system would be free of physical and mental effort. The model posits that if users find a technology easy to use, they will also perceive it as more useful. Therefore, as perceived ease of use increases, it positively impacts perceived usefulness. Research in the context of IS has demonstrated that perceived usefulness and perceived ease of use directly influence behavioral intention, thus eliminating the need for the attitude construct that is central to both TRA and TPB (Davis, 1989).

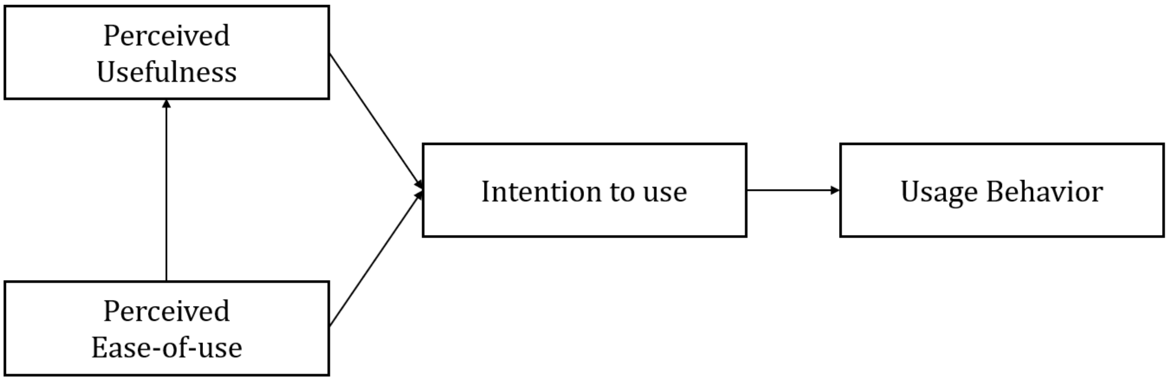

Figure 2-6 – Technology Acceptance Model (Following Davis, 1989)

TAM was later extended into TAM-2 to explain better the factors influencing perceived usefulness, one of the key predictors of behavioral intention to use a technology. TAM-2 incorporated social elements, such as subjective norm, voluntariness, and image, and cognitive elements, such as job relevance, output quality, and result demonstrability, as determinants of perceived usefulness (Venkatesh & Davis, 2000). Subsequently, the Technology Acceptance Model 3 (TAM-3) was adapted to study the determinants of perceived ease of use, the other crucial predictor of behavioral intention to use a technology. TAM-3 introduced four “anchor” variables (computer self-efficacy, perception of external control, computer anxiety, and computer playfulness) and two “adjustment” variables (perceived enjoyment and objective usability) (Venkatesh & Bala, 2008).

These advancements in the TAM framework have provided a comprehensive understanding of the factors that influence user acceptance of technology. TAM was adopted in research, for example, to find out how AI can be more profitable and effective in e-commerce (C. Wang et al., 2023). Also, Xu & Wang (2021) used TAM in relation to the acceptance of robot lawyers. And more recently TAM was utilized regarding the adoption of ChatGPT (Yee et al., 2023).

2.2.4 Unified Theory of Acceptance and Use of Technology

To synthesize existing research on technology acceptance, the Unified Theory of Acceptance and Use of Technology (UTAUT) was proposed in 2003, drawing inspiration from various models, including the Theory of Reasoned Action (TRA) and Technology Acceptance Model (TAM), among others. UTAUT identifies four key constructs—performance expectancy, effort expectancy, social influence, and facilitating conditions—that influence the behavioral intention to use a technology, predicting actual usage (Venkatesh et al., 2003).

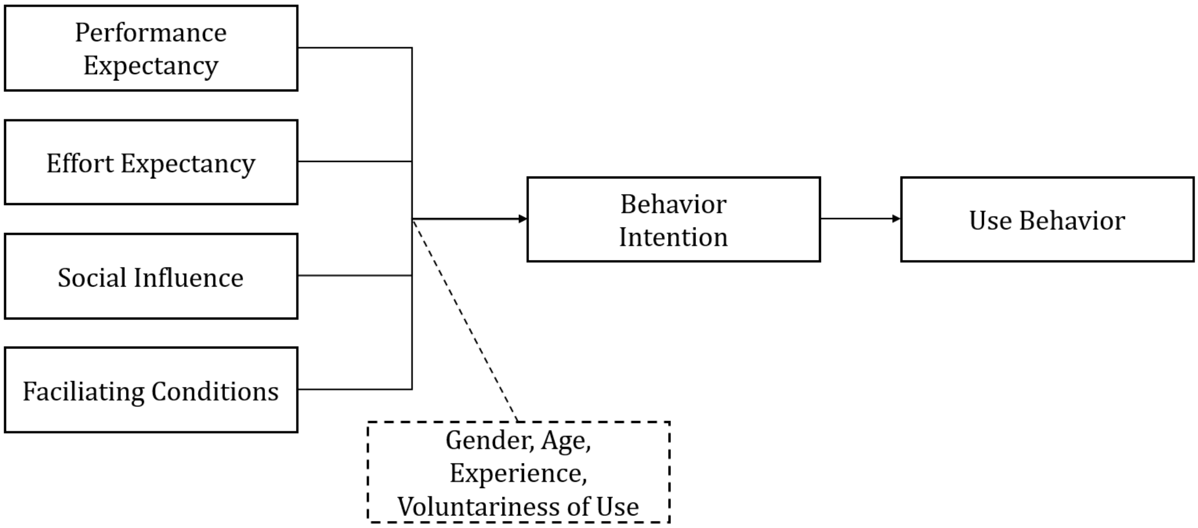

Figure 2-7 – Unified Theory of Acceptance and Use of Technology (Following Venkatesh et al., 2003)

Performance expectancy is defined as the degree to which technology benefits users in performing certain activities. Effort expectancy expresses the extent to which using a technology is free from effort. Social influence pertains to how an individual perceives that important others believe they should use the system. Facilitating conditions refer to the perception of the availability of resources and support to use the technology. Generally, the stronger each of these constructs, the stronger the behavioral intention to accept and use a technology. Additionally, UTAUT includes four moderating variables—age, gender, experience, and voluntariness of use—which are theorized to influence the relationships within the model (Venkatesh et al., 2003).

In 2012, the Unified Theory of Acceptance and Use of Technology 2 (UTAUT-2) was introduced to examine the acceptance and use of technology in a consumer context. UTAUT-2 added three predictors of behavior: hedonic motivation, price value, and habit. Hedonic motivation is the fun or pleasure of using a technology. Price value represents the consumers' cognitive trade-off between the perceived benefits of the technology and the monetary cost of using it. Habit relates to the extent to which individuals tend to perform behaviors automatically due to learning. In this updated model, the voluntariness of use was removed to make the framework applicable in contexts where the use of technology is voluntary, such as in the marketplace (Venkatesh et al., 2012). Further research based on UTAUT-2 extended the model by incorporating an additional independent variable, personal innovativeness in IT, initially tested in an educational context. While some have referred to this extended framework as UTAUT-3, it remains uncertain whether a UTAUT-2 model with one additional variable can formally be called UTAUT-3 (Farooq et al., 2017).

2.2.5 Technology-Organization-Environment Framework

The Technology-Organization-Environment (TOE) framework is one of the most frequently used models to examine technology adoption and its influencing factors (Kinkel et al., 2021). The TOE framework focuses on the organizational level and considers external factors that may impact the organization (Oliveira & Martins, 2011).

The TOE framework identifies three contexts influencing technology adoption within and around an organization: the technological context, the organizational context, and the environmental context (DePietro et al., 1990). The technological context includes internal and external technologies that may be relevant to the firm. It focuses on how an innovative technology can influence the adoption process, encompassing AI, its technical requirements, sub-disciplines, alternative technologies, and the overall enterprise IT infrastructure. The organizational context covers a firm's descriptive measures, such as firm size, managerial structure, and human resources. The environmental context refers to the external environment in which an organization operates, presenting opportunities and constraints for technological innovation (DePietro et al., 1990). The TOE framework is well-supported theoretically and empirically in IS research (Oliveira & Martins, 2011).

In AI adoption research, the TOE framework has been frequently applied. For example, Kruse et al. (2019) utilized the TOE framework to analyze AI adoption challenges in the financial services industry. Kinkel et al. (2021) explicitly concluded that the framework is suitable for studying AI. Other applications include examining AI readiness factors (Jöhnk et al., 2021; Pumplun et al., 2019). Thus, the TOE framework is expected to be useful in organizing findings, interpretations, and revealing connections.

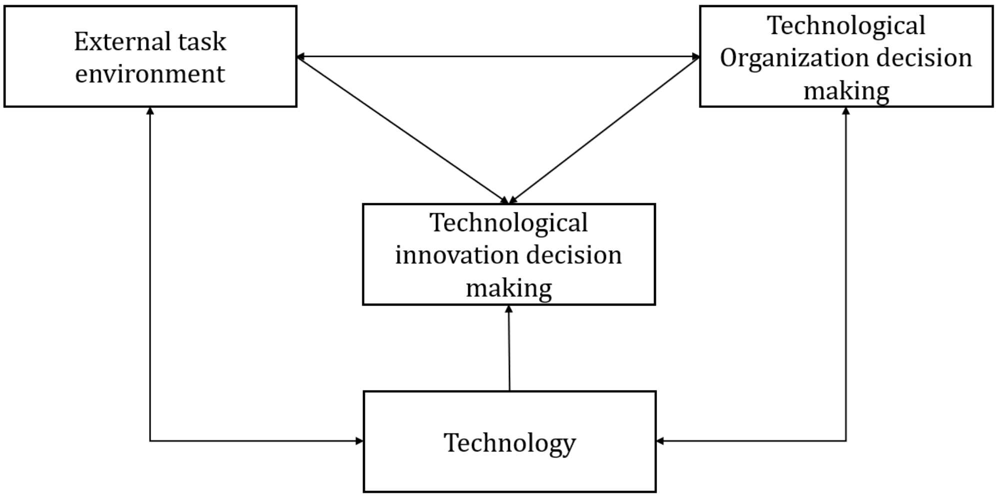

Figure 2-8 – Technology-Organization-Environment (TOE) Framework (Following DePietro et al., 1990)

Additionally, other models have been developed specifically for AI adoption, such as the HACAF model, which directly addresses specific factors of AI adoption identified in the literature based on previous models (Russo, 2024). For instance, research has shown that at the individual level—currently the focus of much research (Agrawal et al., 2024)—factors such as security, trust, intrinsic motivation, price, and social influence play significant roles. Individual factors such as culture also play a crucial role in technology adoption (Chang et al., 2014; Srite & Karahanna, 2006). In contrast, at the organizational level, strategic factors such as top management support, technical competencies, and compatibility with existing development workflows are important (Brynjolfsson & McAfee, 2017; J. Chen et al., 2023; Horani et al., 2023; Radhakrishnan & Chattopadhyay, 2020; Russo, 2024). Conversely, factors like complexity and regulations can hinder the adoption of AI (Horani et al., 2023). Adequate investment in AI is also crucial for sustainably increasing an organization's revenue (Y. S. Lee et al., 2022).

In more detail, the following sections will examine several influencing factors used in this dissertation.

Culture

Several studies emphasize the effect of national cultural dynamics in the organizational context. These studies reveal that various nations manage their organizations differently (Trompenaars & Hampden-Turner, 1997). Different theories have been developed to measure various cultures, with the most well-known being Hofstede's definition of culture as "the collective programming of the mind that distinguishes members of one group [...] from another" (Hofstede, 1991, p. 5). Many cultural theories categorize countries based on one or more dimensions. In the initial version of his theory, Hofstede categorized cultures into four dimensions (Power Distance, Uncertainty Avoidance, Individualism vs. Collectivism, Masculinity vs. Femininity), which later expanded to six dimensions (adding Long Term Orientation and Indulgence vs. Restraint) (Hofstede, 1980). Other theories, such as Fukuyama's distinction between high-trust cultures (where larger organizations can exist) and low-trust cultures (where family is essential) (Fukuyama, 1995) or Hall's differentiation between high context (implicit information and personal relationships) and low context (explicit communication and less importance on context) (Hall, 1976), provide additional perspectives.

Willingness to pay

Another essential factor influencing adoption and acceptance is price. Pricing often depends on willingness to pay (WTP), which describes the maximum amount a person is willing to pay for a product or service (Kalish & Nelson, 1991). At this price, the consumer is indifferent to buying the product (Moorthy et al., 1997). Understanding WTP allows for maximizing sales volume and profit margin, making it a critical concept for pricing (Le Gall-Ely, 2009). However, determining WTP can be challenging, with conjoint analysis emerging as a promising methodology (Breidert et al., 2006). While WTP is often studied at the individual level, it is handled differently within organizations, though there is relevant research on this as well (Goebel et al., 2018).

Bias

Humans often deviate from optimal decisions because it is impossible to always think and act rationally. Several factors contribute to this: people usually cannot process all available information or do not have it; incorrect conclusions can be drawn due to emotional influences or existing prejudices; or heuristics are used for quick decisions (Kahneman, 2012). These influences on rational decisions are known as cognitive biases. Our limited cognitive processing capacity is not the only factor determining our cognitive intelligence. Human cognitive information processing also exhibits systematic distortions, manifested in many cognitive biases (Tversky & Kahneman, 1973, 1974). Cognitive biases are systematic, universally occurring tendencies, inclinations, or dispositions that skew or distort information processes, leading to inaccurate, suboptimal, or erroneous outcomes (e.g., Korteling et al., 2021; Lichtenstein & Slovic, 1971; Tversky & Kahneman, 1981).

In IS research, there is substantial literature on cognitive biases in decision-making (e.g., Hristov et al., 2022; Ni et al., 2019) and biases in the adoption of emerging technologies (e.g., Balakrishnan et al., 2021; Frank et al., 2023; Piehlmaier, 2022). One notable bias is the IKEA effect, which suggests that people value objects more highly if they assemble or create them themselves (Norton et al., 2012). This effect can also influence AI adoption, as proper adoption of AI is essential for its potential to impact an organization (Agrawal et al., 2024; Chuma & De Oliveira, 2023). The interaction between AI and humans is particularly relevant in this context (Henry et al., 2022).

These factors highlight the multifaceted nature of AI adoption, emphasizing the importance of considering both individual and organizational influences.

2.3 Human-AI Collaboration

The research stream on human-AI collaboration investigates the dynamics and outcomes of machines and humans working together (e.g., Boyacı et al., 2023; Fügener et al., 2022). Although AI systems now surpass humans in certain tasks (Jiang et al., 2022), the primary goal is not to replace humans but to enhance collaboration between humans and machines (Jiang et al., 2022; Korteling et al., 2021). Studies have shown that such collaboration can improve performance (Boyacı et al., 2023; Fügener et al., 2021).

One advantage of human-AI interaction is that AI can support humans who do not always make rational decisions, and it can help maintain motivation by allowing task delegation (Fügener et al., 2022). The collaboration can take various forms (Berente et al., 2021). Tasks can be delegated to AI, or AI can delegate and evaluate tasks (Kellogg et al., 2020; Möhlmann et al., 2021).

Additionally, research has demonstrated that when decision outcomes differ between humans and AI, individuals respond differently. For example, experienced physicians often ignore AI advice, whereas novice physicians tend to question their own decisions and are less satisfied with the AI system (Jussupow et al., 2021). To foster effective collaboration, providing training can be beneficial (Korteling et al., 2021). It is also crucial that human knowledge is not lost in the interaction but is actively incorporated into decision-making (Fügener et al., 2021). Therefore, there should be a focus on developing systems that facilitate human-AI collaboration rather than AI making exclusive final decisions (Abel-Karim et al., 2020). This approach is particularly important for maintaining motivation, as humans prefer to engage in challenging tasks (Allan, 2017; Fügener et al., 2022).

2.4 Research Methodologies

In the IS research domain, the accurate implementation and effective presentation of research methodologies are highly valued. A sound research design and its execution are essential for any publication. Given the importance of methodology in IS research, a significant body of literature provides detailed guidance on designing and executing research projects. Five overarching pillars of research methods and their respective guidance have emerged.

Literature Review

Literature reviews are crucial for interpreting and combining existing knowledge (vom Brocke et al., 2009). They aim to uncover relevant sources related to a topic under study, thus contributing to the relevance and rigor of research (vom Brocke et al., 2009). Systematic literature reviews (SLR) are a specific type of literature review, as Okoli and Schabram (2010) highlighted, providing a structured approach to synthesizing existing research. These reviews offer a theoretical background for subsequent research, help understand the breadth of research on a topic, or answer practical questions based on existing studies. Literature reviews are typically part of a research article. However, literature reviews can also constitute original and valuable research on their own, providing a solid starting point for other researchers in the field (Okoli & Schabram, 2010). According to Webster and Watson (2002), a literature search involves querying scholarly databases using keywords and conducting backward and forward searches based on relevant articles. Comprehensive guidelines for performing literature reviews in a structured and reliable manner are provided by scholars such as Okoli and Schabram (2010), Schryen (2015), vom Brocke et al. (2009), and Webster and Watson (2002).

Qualitative Research

Qualitative research methods are essential for investigating various phenomena concerning IS through data from interviews, observations, archival materials, interventions, or design efforts (Conboy et al., 2012). Eisenhardt (1989), Myers and Newman (2007), and Urquhart et al. (2010) provide fundamental frameworks for conducting qualitative research, such as grounded theory, qualitative interviews, and case studies. Grounded theory, distinguished by its continuous interplay between data collection and analysis, aims to develop theory grounded in systematically gathered and analyzed data (Urquhart et al., 2010). Case studies also aim to develop theories, especially for complex social phenomena (Eisenhardt, 1989; Yin, 2014). Qualitative interviews are a critical data-gathering tool across different types of qualitative research, including interpretative, critical, and positivist studies, and are used in case studies, action research, grounded theory studies, and ethnographies (Myers & Newman, 2007). Content analysis is a widely used qualitative research technique for analyzing and extracting findings from qualitative data. It can follow three distinct approaches: conventional, directed, or summative, each interpreting meaning from text data while adhering to the naturalistic paradigm (Hsieh & Shannon, 2005).

Design Science Research (DSR)

Design Science Research (DSR) in IS involves constructing socio-technical artifacts such as decision support systems, modeling tools, governance strategies, methods for IS evaluation, and IS change interventions (Gregor & Hevner, 2013). The design-science paradigm achieves knowledge and understanding of a problem domain and its solution by building and applying designed artifacts (Hevner et al., 2004). While developing strong theory is a significant contribution of DSR, developing partial or incomplete theories or empirical generalizations in the form of new design artifacts is also valuable (Gregor, 2006). Frameworks and step-by-step guidelines can be found in (e.g., Hevner et al., 2004; Kuechler & Vaishnavi, 2008; Peffers et al., 2007).

Mixed Methods Research

Mixed methods research combines quantitative and qualitative research methods within the same inquiry to develop rich insights into phenomena that cannot be fully understood using only one method (Venkatesh et al., 2013). Although the terms "mixed methods" and "multimethod" are often used interchangeably, there are significant conceptual differences between the two (see Venkatesh et al., 2013). Mixed methods research uses concurrently or sequentially quantitative and qualitative approaches to understand a phenomenon of interest. This approach values both quantitative and qualitative perspectives to develop a deep understanding. For instance, a researcher might use interviews and surveys to collect data about a new IS implementation or employ ethnography and field experiments to understand the same phenomenon. Mixed methods design strategies provide IS researchers with a powerful mechanism to address rapidly changing environments where existing theories and findings may not sufficiently explain a phenomenon (Venkatesh et al., 2013).

Quantitative Research

Quantitative research methods, such as surveys, experiments, and simulations, are integral to IS research. Often administered to broad samples, surveys result in generalizable conclusions using structured questionnaires (W. R. King, 2005; Mazaheri et al., 2020). They can be categorized into cross-sectional and longitudinal surveys, with the former involving one-time data collection and the latter surveying the same individuals multiple times to measure changes over time (Moorman, 2008; Srivastava et al., 2015). Surveys have been a key element of IS research for decades, allowing researchers to observe phenomena in broad samples and validate relationships statistically (Dinev et al., 2013; Melnyk et al., 2012; Newsted et al., 1998; Recker, 2021). Simulations, such as agent-based modeling (ABM), involve simulating large numbers of autonomous agents interacting with each other and their environment to observe emergent patterns (E. R. Smith & Conrey, 2007). Experiments are also a vital quantitative method for testing hypotheses and examining causal relationships. Here, often one or more variables are varied and thus manipulated which can be used as controls to derive relations and influences (Kirk, 2009; Novikov & Novikov, 2013).

By employing these diverse research methodologies, IS researchers can enhance the robustness and validity of their studies, contributing valuable insights to the field.

3 Paper A: Exploring the Effect of National Culture on Emerging Technologies: A Glimpse Into the Future

Title

Exploring the Effect of National Culture on Emerging Technologies: A Glimpse into the Future

Authors

Mehler, Maren F., Turan Akdag, Merve, and Zöll, Anne

Publication Outlet

Proceedings of the Pacific Asia Conference on Information Systems (PACIS 2023), Nanchang, China

Abstract

As organizations become increasingly globalized, understanding the impact of national culture on the successful usage and adoption of emerging technologies is crucial. National culture influences the strategies of organizations, for instance how the employees successfully adopt and use emerging technologies. While the effect of national culture has been widely observed in information systems, it is still challenging to measure the influence of national culture on the usage and adoption of emerging technologies. To contribute to the existing body of knowledge, we conducted a structured literature review on how previous work measured national culture to provide a starting point for further theory development. For instance, our findings emphasize the need to measure culture at an individual level. Finally, we developed a research agenda to provide a starting point for developing theories to measure the influence of national culture on emerging technologies.

Keywords

National Culture, Emerging Technologies, Research Agenda, Cultural Theories

3.1 Introduction

In recent decades, economic interdependencies have increased sharply worldwide. The scarcity of resources during the pandemic exemplifies how strong the international interdependencies are (OECD, 2020). International stakeholders operate across national borders and rely on many different cultures working together to ensure the organization's success. Employees of various national cultures collaborate between different organizations. Expanding product and digital service offerings to an international market can greatly benefit organizations by increasing their potential sales market and enhancing overall organizational success (Wulf et al., 2017). At the same time, emerging technologies can support organizations in cross-border collaboration and sales. To effectively cater to customers from diverse cultural backgrounds, organizations must possess a comprehensive understanding of cultural differences (Moser & Deichmann, 2021). A deep cultural understanding plays a pivotal role in fostering teamwork for successful adoption of emerging technologies, as well as in retaining and attracting international customers, both of which are indispensable for organizational success (e.g., Fang et al., 2011).

The challenges that emerge when crossing cultural boundaries primarily pertain to the diminished effectiveness of management processes. This issue arises when a multinational organization disregards national cultural differences and attempts to apply management practices developed by subsidiaries in their own cultural context, which have already proven effective within that specific culture. Recognizing and comprehending cultural variations is critical and it must be acknowledged that due to these differences, there is no one-size-fits-all approach for running an organization and adopting emerging technologies to push innovation forward (Trompenaars & Woolliams, 2003). Over the past decades, there has been a growing interest in Information Systems (IS) research literature on the impact of national cultural differences in IS research field (Myers & Tan, 2002). Previous studies have investigated the characteristics of various cultures that might systematically influence whether and for what reasons nations adopt available emerging technologies (e.g., D. J. Kim et al., 2016; Shore & Venkatachalam, 1996). In general, these studies agree that national culture is important as understanding culture and its impact can lead to a successful deployment of emerging technologies in a global environment (Ives & Jarvenpaa, 1991; Myers & Tan, 2002; Shore & Venkatachalam, 1996). To capture differences between national cultures, many authors have already investigated the effect of national culture and developed measurement instruments for national culture (e.g., Fukuyama, 1995; Hall & Hall, 1990; Hofstede, 2011). However, there is ongoing criticism regarding the application of Hofstede's dimensions at both the individual and organizational levels, as these dimensions primarily reflect country-level analysis and fail to fully explain individual behavior (Srite & Karahanna, 2006; Venaik & Brewer, 2013). Furthermore, Hofstede's concept of national culture assumes a correlation between cultural differences and the territorial boundaries of a nation, which seems unrealistic given that within a single nation or organization, many individuals from different national cultures work together (Myers & Tan, 2002). Moreover, the notion that each nation has its own national culture considered is inherently not correct, and nations continue to change in their form and makeup, which means national culture has a dynamic component (Myers & Tan, 2002). Thus, the call of researchers is related to developing new theories and measurements in the field of the effect of national culture on emerging technologies (Myers & Tan, 2002). To take the first steps in this direction, we examine how previous literature has defined and measured the impact of national culture, and we aim to answer the following research question (RQ): *How previous research has measured the influence* *of national culture on emerging technologies, and what promising avenues exist for future research in this field?*

To answer this RQ, we conducted a structured literature review according to vom Brocke et al. (2009). This methodology allowed us to provide an overview of existing operationalization and to identify future research steps. To the best of our knowledge, Chu et al. (2019) are the only ones that dedicated their research on a structured literature review related to the measurement of national culture in IS. However, our study differs regarding three aspects: First, we considered recent work from the time period 2000-2022 (instead of 2006-2016). Second, we focused on the adoption and usage of emerging technologies in organizations, while Chu et al. (2019) examine general IS literature. Third, we focused on the effect of national culture and examined papers that adopted national culture as a moderator. Chu et al. (2019) also analyzed papers that address the influence of culture and considered national culture as an independent variable. However, according to Baron and Kenny (1986) the manipulation through personal or situational variables such as age or gender should be measured as moderators. Given that culture is inherently individualistic (Srite & Karahanna, 2006), our interest lies in comprehending the extent to which the connection between independent and dependent variables can be amplified or diminished by the moderating influence of culture.

We contribute to research by providing a starting point for developing new theories in the field of national culture by first presenting a concept matrix of how the effect of national culture has been measured in IS research. In this matrix, we highlight which theories have been used in national cultural research. Second, we present a research agenda aimed at advancing the understanding of the influence of national culture on emerging technologies, thus paving the way for future research in this domain.

The rest of this paper is structured as follows: Section 2 outlines the methodology chapter. We present the methodology early in the paper since the different identified theories in the structured literature review are also part of the results. In the third chapter, we describe the theoretical background. After outlining the most commonly used theories, we show the findings of our structured literature research in chapter four. Subsequently, we present the research agenda and conclude the paper with the contribution, limitations, and future research.

3.2 Methodology

To answer the RQ on how the effect of national culture was measured in IS research and influenced the adoption of emerging technologies, we conducted a structured literature review according to the approach of vom Brocke et al. (2009). As shown in Figure 1, this framework suggests five steps: 1) definition of review scope, 2) conceptualization of the topic, 3) literature search, 4) analysis and synthesis, and 5) research agenda.

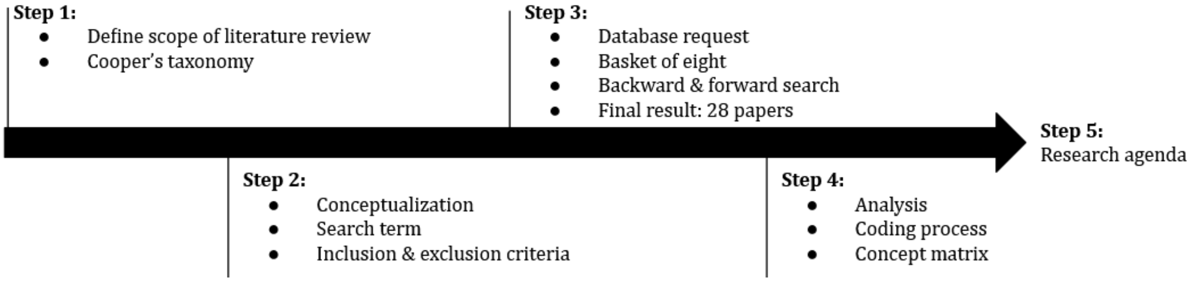

Figure 3-1 – Approach According to vom Brocke et al. (2009)

In **step 1**, we defined the scope and set up goals according to Cooper's taxonomy, as shown in Table 1 (Cooper, 1988). The taxonomy consists of six characteristics that provide researchers with general guidance for a successful structured literature review. The categories comprise focus, goal, organization, perspective, audience, and coverage. We discussed the *focus* at the beginning of the research project, which primarily examines the effect of national culture on the adoption and usage of emerging technologies. In particular, we decided to concentrate on papers which used national culture as a moderator as culture is an individual trade (Srite & Karahanna, 2006) influencing existing relationships between independent and dependent variables, which should be measured as a moderator according to Baron and Kenny (1986). Therefore, out of scope are papers that consider culture as, e.g., an independent variable. The *goal* of this structured literature review is to develop a research agenda and pave the way for further research. Our paper is *organized* conceptually since we aim to develop a concept matrix that structures the previous literature according to the recommendation of vom Brocke et al. (2009). In general, we took a neutral *perspective*. The *audience* is either general researchers or scholars specializing in the impact of national culture on the adoption and usage of emerging technologies. Finally, we chose a representative *coverage*, i.e., we provide individual examples from literature to represent a larger amount of topics.

Characteristics	Categories			
Focus	Research outcome	Research methods	Theories	Applications
Goal	Integration	Criticism	Central issues	
Organization	Historical	Conceptual	Methodological	
Perspective	Neutral representation		Espousal of position	
Audience	Specialized scholars	General scholars	Practitioners	General public
Coverage	Exhaustive	Exhaustive/ selective	Representative	Central/pivotal

Table 3-1 – Taxonomy According to Cooper (1988)

In **step 2**, we conceptualized the literature review. Since our aim was to provide broad coverage of the previous literature on how the influence of national culture on emerging technologies is measured, we agreed on the general search term *"culture" OR "cultural"*. Since we focused on measuring the effect of culture on emerging technologies, we defined the following inclusion criteria: i) The paper examined national culture in organizations, ii) the paper compared different countries concerning their effect of national cultures on the adoption and usage of emerging technologies, iii) national culture was designed as a moderator, and iv) to ensure representative coverage of high-quality papers, we considered papers of the basket of eight journals. In addition, we defined the following exclusion criteria: i) We do not consider papers published before 2000 to focus on the more recent literature, ii) we do not consider papers where culture was designed as an independent variable, and iii) we exclude papers with a focus on organizational culture since our research subject concentrate on national culture.

In **step 3**, we queried the databases Web of Science and specifically considered the basket of eight journals: European Journal of Information Systems, Information Systems Journal, Information Systems Research, Journal of the Association for Information Systems, Journal of Information Technology, Journal of Management Information Systems, Journal of Strategic Information Systems, and MIS Quarterly. The request revealed 279 results. We scanned the titles and abstracts of the papers and applied the inclusion and exclusion criteria. As a result, a total number of 72 papers remained. Afterward, we read these papers carefully and applied the inclusion and exclusion criteria once again. In this step, we focused particularly on the papers in which the national culture was considered as a moderator variable, of which 24 papers remained. Backward and forward search were then carried out. The backward search revealed no relevant paper. However, the forward search identified four more papers which fulfilled the inclusion and exclusion criteria. The final selection consists of 28 papers which serve as the basis for further analysis.

In **step 4**, we analyzed the papers and identified the used measurements for national culture. In particular, we classified the papers related to the following dimensions: application scenarios, culture-related theories and frameworks, data collection, data analysis, and the level of measuring culture. Concerning the research topic, we used VOSviewer to identify the relevant research contexts. VOSviewer is a tool to visualize the network of keywords based on bibliographic data (VOSViewer, 2023). We excluded keywords on national cultural differences such as “national culture”, “united-stated”, or “uncertainty avoidance” as those focus on the overall area of our literature review.

In general, we coded the papers based on these dimensions and subordinary topics according to the recommendations of Schryen (2015). In particular, we followed the approach of vom Brocke et al. (2009) and developed a concept matrix considering these extracted dimensions. In this vein, we identified embedded national cultural theories that measure national culture. The three main theories are Hofstede’s Cultural Dimensions Theory, Fukuyama’s Theory of Trust and Social Capital, and Hall’s Context Theory (Fukuyama, 1995; Hall, 1976; Hofstede, 1980). Based on these theories, we developed a research agenda in **step 5**.

3.3 Theoretical Foundation and Cultural Theories

After outlining our methodology, we present the theoretical foundation on emerging technologies and especially on the measurement of national culture, which is also part of our concept matrix.

3.3.1 Emerging Technologies

Emerging technologies encompass technologies that are presently undergoing development or reaching a stage of maturation (Wheatley & Wilemon, 1999). As this definition is contingent upon the specific time period being considered, it will encompass distinct technologies accordingly. For instance, artificial intelligence (AI), particularly ChatGPT and the metaverse, are presently regarded as emerging technologies (Gartner, 2022). In recent years, notable examples of emerging technologies include autonomous driving, the Internet of Things, and wearables

(Gartner, 2015). To gain a competitive edge, organizations must invest in these novel technologies and effectively harness their potential. In the literature, the term “emerging technologies” is frequently utilized; however, it is seldom defined. For example, Conger et al. (2013) examined the impact of emerging technologies on privacy, and Bhattacherjee (1998) addressed the management of emerging technologies. Teams explore the potential of emerging technologies by developing proofs-of-concept and proof-of-value prototypes (Nunamaker Jr. et al., 2015). In doing so, they try to identify promising use cases for emerging technologies. Many organizations become successful by selling their products and services internationally. The products and services can be emerging technologies, but also products and services that have been manufactured using these technologies.

However, there are also challenges concerning the usage and adoption of emerging technologies across different cultures. The acceptance of technologies and management or leadership approaches can vary across different countries due to cultural differences, socioeconomic factors, and regulatory environments (Chang et al., 2014; Srite & Karahanna, 2006). Each country has its own unique cultural norms, values, and beliefs that shape people’s attitudes and preferences towards technology. As a result, it cannot be assumed that high acceptability in one country translates to high acceptability in another. Socioeconomic factors, such as income levels, education, and infrastructure, significantly influence technology acceptance. Countries with higher levels of development and greater access to resources generally exhibit higher rates of acceptance and adoption compared to countries facing limited resources or infrastructure challenges. Additionally, the regulatory environment plays a crucial role in shaping the adoption of emerging technologies. For example, European countries have distinct data protection rules compared to the United States, which leads to different initial conditions. Consequently, the field of IS has developed a keen interest in measuring the impact of national culture to address these challenges (Myers & Tan, 2002).

The rapid advancement of AI represents an excellent example that brings forth challenges in understanding the cultural implications related to the implementation and adoption of self-learning algorithms. Awad et al. (2018) introduced an online experimental platform called Moral Machine. In this study, cross-cultural differences have been observed regarding ethical preferences. Particularly when AI algorithms are trained on a database from one specific culture and subsequently applied in a different culture, it can result in decision support that is incompatible with the target culture. To identify such disparities and cultural nuances, cultural studies play a vital role. These findings emphasize the importance for AI developers to pay attention to the general ethical preferences in different countries.

3.3.2 National Culture

Several studies emphasize the effect of national cultural dynamics in the organizational context. These studies reveal that various nations manage their organization differently (Trompenaars & Hampden-Turner, 1997). Even though scholars agree on the relevance of national culture in organizational research, there is still no clear consensus on which cultural theory is superior (Hofstede, 2001; Javidan et al., 2006; Venaik & Brewer, 2013). In the following, we present an overview of the identified theories in Table 2.

Theo-ries	Hofstede's Cultural Dimensions Theory	Fukuyama's Theory of Trust and Social Capital	Hall's Context Theory
Overview	Defines culture as "the collective programming of the mind that distinguishes members of one group […] from another."	Focuses on the influence of social capital on economic and social development, which overall includes trust.	Context dimension explains behavior in intercultural communication.
Dimensions	**Four original dimensions:** Power Distance, Uncertainty Avoidance, Individualism vs. Collectivism, Masculinity vs. Femininity. **Additional dimensions:** Long Term Orientation, Indulgence vs. (Self-)Restraint	**One dimension:** High trust culture (larger organizations can exist) vs. Low trust culture (family is essential)	**One dimension:** High context (implicit information and personal relationships) vs. Low context (communication is explicit, and context is less important)

Table 3-2 – Cultural-Related Theories

3.3.2.1 Hofstede's Cultural Dimensions Theory

One of the best-known theories of culture was developed by Geert Hofstede. From 1967 to 1973, Hofstede created his "cultural dimensions" theory based on an empirical study with more than 110,000 IBM employees using factor analysis (Hofstede, 2011). He defines culture as "the collective programming of the mind that distinguishes members of one group or category of people from another" (Hofstede, 1991, p. 5). His theory explains culture on a national level and thus looks at cultural differences between different nations (Hofstede, 1980). The dimensions are measurable aspects of culture that can be compared across nations. The scores normally range from 0 to 100. However, later-added nations can also exceed the score of 100. The scores represent a relative and not an absolute position, meaning they are only measures of differences (Hofstede, 2001).

The first version of Hofstede's Cultural Dimensions Theory, published in 1980, consists of four cultural dimensions: "Power Distance Index (PDI)", "Uncertainty Avoidance Index (UAI)", "Individualism vs. Collectivism (IDV)", and "Masculinity vs. Femininity (MAS)" (Hofstede, 1980). The *Power Distance (PDI)* dimension provides information about the degree of inequality in power relations between superiors and subordinates. It indicates the extent to which the less powerful expect power to be distributed unequally. People from cultures with higher power distance expect some degree of inequality in terms of power and wealth. The lower a nation's scores on the scale, the more stable the cultural environment is expected to be and the higher the expected cooperative interaction between different power levels. At high power distance, hierarchy is tolerated and seen as legitimate. Low power distance indicates a desire for equality and justification (Hofstede, 1980). *Uncertainty Avoidance Index (UAI)* explains whether an uncertain future is generally seen as negative with a desire for more rules and security (high value) or whether it is relatively quickly accepted and thus increases the willingness to take risks. This dimension is related to the need for security. People from nations with high uncertainty avoidance try to avoid new and unknown situations (Hofstede, 1980, 2011). Another dimension, *Individualism vs. Collectivism (IDV)*, defines to what extent individual self-determination or collectivistic integration is more critical. Nations that score high on this dimension are individualistically oriented. This means that people are self-oriented, constantly seek their own advantage, and put their needs above those of the group. In contrast, collectivist-oriented people pay more attention to the wishes of the group and the opinions of other group members (Hofstede, 1991). They consider it very important to help and share within the group (Hofstede, 1980).

Finally, a distinction is made between *Masculinity and Femininity (MAS)*, which should not be confused with gender, as only the characteristics assigned to men or women are meant. Nations that achieve a high score in this dimension have more masculine characteristics. This means that people focus on work and favor earnings and promotion. On the other hand, people from feminine cultures value caring and personal goals, such as a friendly environment (Hofstede, 1991). The fifth cultural dimension, *Long Term Orientation (LTO)*, was added to Hofstede's Cultural Dimensions Theory in 1991 (Hofstede, 2011). This dimension refers to the extent to which societies are geared towards focusing more on short-term success or long-term solutions. In long-term cultures, the focus is on building long-term personal relationships or relationship networks. Furthermore, great respect is shown for persons of higher rank and older people and traditions. Contrary, short-term cultures prefer short-term wins and have guidelines about what is good and bad (Hofstede, 2001). The sixth cultural dimension of Hofstede, which was added to the theory in 2010, is *Indulgence vs. (Self-)Restraint (IVR)* and is based on the research of Michael Minkov (Hofstede, 2011). Nations in the upper score attach great importance to joie de vivre and fun, i.e., enjoying life. They allow an instant satisfaction of basic human needs. Self-restraint means that the satisfaction of people's needs is controlled by society and regulated by strict norms (Hofstede, 2011).

3.3.2.2 Fukuyama's Theory of Trust and Social Capital

Another national culture theory came from Francis Fukuyama in 1995, who considered cultural differences on a national level. Fukuyama studied the influence of a society's social capital on the development of its economy. Social capital includes different characteristics of a society that improve its efficiency by facilitating collaboration. Above all, this contains mutual trust. Trust arises when people share the same moral values. It is the basis for a society's economic and social development. In organizations, there are so-called transaction costs between organizations. According to Fukuyama, organizations with a high level of trust can work together more efficiently because, for example, transaction costs are low since economic actors trust each other and do not need detailed contracts. On the other hand, a low level of trust in a society makes market activities more difficult because transaction costs increase. Therefore, trust and social capital are essential prerequisites for a functioning economy and ensure social cohesion. In his theory, Fukuyama divided the nations into "high trust" and "low trust" groups. According to his assessment, family is essential in "low-trust cultures" as people can only trust their family and have a low level of trust towards others. Therefore, family organizations are predominant in low-trust cultures. In "high-trust cultures", on the other hand, larger organizations exist. Fukuyama justifies this with his concept of spontaneous sociability, which is based on trust between people and means organizations can grow beyond the family (Fukuyama, 1995).

3.3.2.3 Hall's Context Theory

The context dimension in Edward T. Hall's Context Theory from 1976 is often used to explain behavior in intercultural communication (Hall & Hall, 1990). Like Hofstede and Fukuyama, Hall's Context Theory describes the culture on a national level (Hall, 1976). The context dimension represents the information surrounding an event or conversation that is important to the meaning of that event (Hall & Hall, 1990). The term "context culture" describes how people in a culture treat each other or, for example, interact with each other in a conversation (M. S. Lee et al., 2007). Hall differentiated between "high context" and "low context" and classified the nations based on his criteria. In a "high-context culture", there is little explicit information in a conversation since most of the relevant information is in the context of the conversation or has already been internalized by the people. This makes the communication more personal, as the conversations are kept very simple but often have more profound, implicit meanings. Close personal relationships are essential in this culture. In a "low-context culture", on the other hand, the information is explicitly communicated in a conversation. People are not closely related, and communication is more impersonal (Hall, 1976). Thus, people need to pay attention to the context of a conversation to understand the message. This is not the case in a low-context culture because the exchange of information is explicit. For this reason, people in a "high-context culture" do not expect any background information as they are always informed, especially about the people who are important to them in life. On the other hand, people from a "low-context culture" need this background information to understand what is happening. Since "high-context cultures" feel irritated by too much information and "low-context cultures" are at a loss when they do not get enough information, it makes intercultural communication challenging (Hall & Hall, 1990).

3.4 Findings of the Structured Literature Review

The 28 papers from the structured literature review, that measure the effect of national culture in IS research, are shown in Table 4. These are analyzed in more detail in this section by providing a context matrix consisting of their topics, culture-related theories, data collection, data analysis, and the level of measuring culture. Additionally, the papers provide details on the nations compared, along with the publication years and journals in which they were published. Interestingly, every paper except M. Zhang et al. (2013) found national cultural differences.

Applications scenarios: Due to the heterogeneity of the papers, we used VOSviewer to extract the application scenarios. Figure 2 illustrates four categories regarding "application scenarios": Management of organizational processes, privacy, technology adoption, and e-commerce (colors correspond to the papers in Table 4).

- Management of organizational process: We identified twelve papers focusing on managing processes, including keywords such as software project management, knowledge sharing, or decision-making. Emerging technologies within this category are either used in organizational processes to improve or implement them. The effect of national culture is examined, for example, by Keil et al. (2000) on the influence of sunk costs on the decision to continue software projects. The findings indicate that individuals within a culture characterized by a low uncertainty avoidance score exhibit a higher tendency to persist with a project. In addition, scholars identified a difference between Chinese and American employees regarding the motivation for knowledge sharing (Chang et al., 2014). In contrast to American employees, Chinese employees can be motivated by rewards. M. Zhang et al. (2013) investigated the effect of culture on the use of IT capability—including emerging technologies—on export performance and found no difference between Chinese and American "born-global" organizations.
- Privacy: The privacy category includes five papers that have a common focus on privacy, in particular on the Internet, information privacy, and technology. On the one hand, various emerging technologies must fulfill privacy requirements (Conger et al., 2013). On the other hand, a diverse array of emerging technologies is employed to attain privacy. For example, Dinev et al. (2009) examined the national cultural difference in user attitudes and behavior toward technologies that are supposed to protect individuals against attacks. They find a difference between the US and South Korea, recommending that such emerging technologies should be designed depending on the national culture. Kim et al. (2016) investigated the

impact of web assurance seal services on trust in e-commerce and, in particular, found an effect of these services in the US but not in South Korea.

- Technology adoption: The category technology adoption comprises four papers dealing with the adoption of emerging technologies that also relate to user acceptance. Both McCoy et al. (2007) and Srite and Karahanna (2006) found an effect of national culture on technology adoption using the Technology Acceptance Model.
- E-commerce: The last category deals with e-commerce, trust, and behavior. Cyr (2008) concluded that the design of websites is influenced by national culture. Sia et al. (2009) identified that depending on the national culture, different web strategies should be used to influence the buying behavior of customers.

Overall, an effect of national culture is demonstrated in the adoption, implementation, and use of emerging technologies in all four categories.

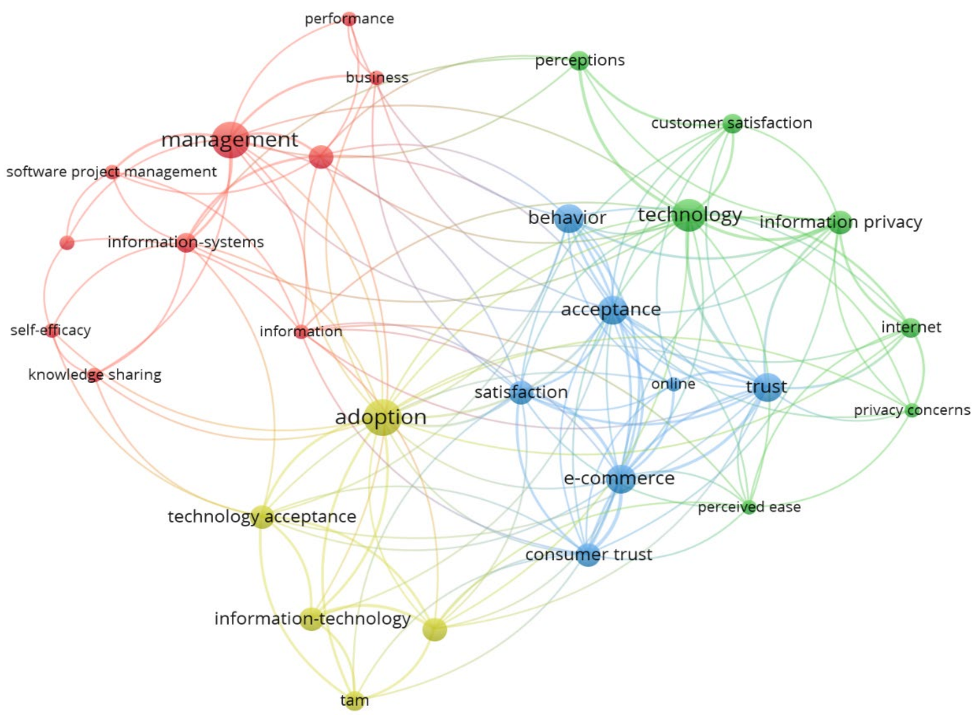

Figure 3-2 – Network of Co-occurrence of Keywords

Culture-related theories: In the next part, we present the used “culture-related theories” which we already described in detail in the chapter theoretical foundations and cultural theories. Our first finding is that Hofstede’s Cultural Dimensions Theory is the most prominent in recent years. They are applied in 23 out of 28 papers. It is worth noting that despite the existence of six distinct dimensions in Hofstede’s Cultural Dimensions Theory we identified only one paper that incorporates all of them, including the sixth dimension of *Indulgence vs. Restraint* (Guan et al., 2022). Only three papers referred to Hofstede’s Cultural Dimensions Theory with six dimensions (Guan et al., 2022; Moser & Deichmann, 2021; Z. Zhou et al., 2015). Twelve papers referenced Hofstede’s Cultural Dimensions Theory with five dimensions, and six of those included all five dimensions. Another nine papers quoted the theory with four dimensions and only four used all of them. Thus, only 11 out of 23 papers that applied Hofstede presented the effect of all cultural dimensions. The dimensions *Individualism vs. Collectivism* and *Uncertainty Avoidance* were used most frequently, a total of 21 and 18 times, respectively, and *Indulgence vs. Restraint* only twice.

Hofstede's Cultural Dimensions Theory offers several advantages: It is applicable at the national level, providing a framework to understand cultural differences across countries (Tuunanen & Kuo, 2015). Moreover, it considers multiple dimensions simultaneously, enabling a comprehensive analysis of cultural variations. However, there are significant disadvantages associated with applying Hofstede's Cultural Dimensions Theory at the individual and organizational levels, as highlighted by Srite and Karahanna (2006). The theory overlooks the existence of cultural and ethnic groups that transcend national boundaries, failing to account for diversity within nations. Additionally, the theory relies on the assumption that cultural differences align strictly with the territorial boundaries of nation-states, which becomes problematic as the concept of nation-states itself is a recent phenomenon. Finally, researchers have found that the relationship between national cultural values and work-related values and attitudes is intricate and not fully explained by Hofstede's model (Myers & Tan, 2002).

Several other theories were applied in the papers. Fang et al. (2011) employed the GLOBE extension proposed by House et al. (2004). In this theory, Hofstede's original dimensions are divided into nine dimensions *In-group Collectivism, Uncertainty Avoidance, Institutional Collectivism, Gender Egalitarianism, Humane Orientation, Power Distance, Performance Orientation, Future Orientation, and Assertiveness* by a collaboration of numerous researchers. The Globe study is an extension of Hofstede's original cultural dimensions framework. The advantages of the Globe study include the inclusion of perspectives from various researchers and companies, which increases its reliability. However, it is important to note that the Globe study is still fundamentally based on Hofstede's work, which means it shares both the advantages and disadvantages of his original framework.

Dinev et al. (2006) used Fukuyama's Theory of Trust and Social Capital, where nations are divided into "high-trust cultures" and "low-trust cultures", depending on the level of trust between humans outside of their own family (Fukuyama, 1995). Fukuyama's Theory of Trust and Social Capital highlights the significance of social capital for the effective functioning of modern economies and the stability of liberal democracies. Social capital serves as the cultural foundation of contemporary societies, contributing to various aspects of societal dynamics (Myers & Tan, 2002). However, a disadvantage of Fukuyama's Theory of Trust and Social Capital is its single-dimensional approach, which may constrain its ability to provide a comprehensive understanding of the intricate dynamics of trust and social capital (Myers & Tan, 2002).

Kim (2008) and Kim et al. (2016) both used the theory of context according to Hall (1976). Hall's Context Theory has both advantages and disadvantages in understanding cultural dynamics. One advantage of Hall's Context Theory is that it provides a theoretical foundation for researchers, as demonstrated by Afrouzi (2021) analyzing the influence of personal context culture on humanitarian behavior intention in different cultural contexts. However, a notable disadvantage of Hall's Context Theory is its classification of cultures based on a single dimension (Myers & Tan, 2002; Tuunanen & Kuo, 2015). This limited perspective may overlook the multidimensionality of cultures and the complexities involved. Furthermore, Hall's Context Theory is primarily applicable to specific domains, particularly in relation to the context surrounding events or conversations. This narrow focus restricts its broader applicability in understanding cultural variations across different aspects of society.

Last, Keil et al. (2007) included a Chinese concept dividing people into one of the two categories of "LIAN" and "MIANZI" which refers to face-saving. LIAN describes that a person is bound to rules of conduct and MIANZI that a person intends to keep a position in a social hierarchy (Earley, 1997). Also, some papers, such as Tuunanen and Kuo (2015) and M. Zhang et al. (2013) compared two or more nations with each other and used no theory at all. Interestingly, if two theories are considered, the Hofstede's Cultural Dimensions Theory is always used as a baseline theory.

Data collection: Furthermore, it is worth examining the methodologies used in the papers to understand what type of research has already been performed to measure the effect of national culture on emerging technologies. We generally identified the following data collection methods: Analysis of user-generated content, Delphi method, interviews, experiments, and surveys. Two papers derived insights from user-generated content (Guan et al., 2022; Yang et al., 2018) and one paper used a Delphi method (Schmidt et al., 2001). Two papers conducted interviews (Miltgen & Peyrat-Guillard, 2014; Tuunanen & Kuo, 2015), and overall, six experiments were carried out (e.g., Keil et al., 2000; Sia et al., 2009), including the two papers that also performed a survey (Cyr, 2008; Cyr et al., 2009). Most papers (19 out of 28) conducted a survey (e.g., Davison et al., 2009; Moser & Deichmann, 2021). In terms of survey participants, students were included as respondents in nine instances (e.g., Dinev et al., 2009; D. J. Kim et al., 2016). Furthermore, the surveys included a substantial number of individuals who are actively engaged in communities, such as social networks (e.g., Z. Zhou et al., 2015). The third large group of participants is IT professionals or managers (e.g., Davison et al., 2009; M. Zhang et al., 2013). Overall, the diversity of data collection methods shows that the data to measure the influence of national culture varies. On one hand, this approach offers the advantage of providing a more thorough and comprehensive overview. On the other hand, it poses challenges in accurately measuring the impact of national culture, making it a complex and nontrivial task.

Data analysis: Papers that adopted the Delphi method or interviews as methodology present national cultural differences individually for each nation. For example, Tuunanen and Kuo (2015) counts the codes extracted from the interviews and provides the frequency per nation descriptively.

The data analysis methods for surveys are more diverse. In general, we identified three approaches which are based on partial least square modeling (PLS) (e.g., Chang et al., 2014; Dinev et al., 2006; D. J. Kim et al., 2016). First, culture was multiplied by the respective variables as a moderator, resulting in an interaction term (e.g., Hoehle et al., 2015; Vance et al., 2008). Second, a multigroup analysis is performed to examine the significant difference between two or more national cultures (e.g., D. J. Kim, 2008; D. J. Kim et al., 2016; Sia et al., 2009). Third, they calculated the regression separately of the different national cultures and compared the results (e.g., the significances of the path coefficients) (e.g., Moser & Deichmann, 2021; Srite & Karahanna, 2006).

Level: We identified two different types of measuring culture. First, on a national level, and second on an individual level. Historically, the measurement of culture has evolved through the development of Hofstede's dimensions, leading to a rating of each nation's scores on the six dimensions of Hofstede. Scholars used these dimensions to prove national cultural differences in their models (e.g., Sia et al., 2009; M. Zhang et al., 2013). The responses from one nation were grouped and evaluated together and compared with those from the other nations. However, Srite and Karahanna (2006) questioned the measuring of the effect of national culture on the acceptance of emerging technologies on a national level, as adopting technology is a personal construct, and individuals might not identify with a given culture. Consequently, they adopted an individual-level measurement of the dimensions from Hofstede's Cultural Dimensions Theory, marking the first instance of this approach. However, six papers have followed them (Chang et al., 2014; Yan Chen & Zahedi, 2016; George et al., 2018; Hoehle et al., 2015; McCoy et al., 2007; S. Sharma et al., 2022). The majority continue to adopt culture as a national construct.

Additional findings: Due to space limitations, we did not present the following findings in the concept matrix. Thus, a review of the nations observed in the 28 papers reveals that, on the one hand, often two nations were compared with each other, and, on the other hand, the United States was the most frequently compared nation. We present these findings in more detail in Figures 3 and 4 such as other scholars (e.g., Chu et al., 2019). Figure 3 demonstrates that more than half of the papers compare two nations (e.g., Dinev et al., 2006; Keil et al., 2007). Six papers compared three nations (e.g., Cyr et al., 2009; Schmidt et al., 2001). One paper four nations (Hoehle et al., 2015) and the remaining four papers seven (Miltgen & Peyrat-Guillard, 2014), 24 (McCoy et al., 2007), 30 (Srite & Karahanna, 2006), 148 (Guan et al., 2022) and two papers did not specify the number of nations (S. Sharma et al., 2022; Z. Zhou et al., 2015).

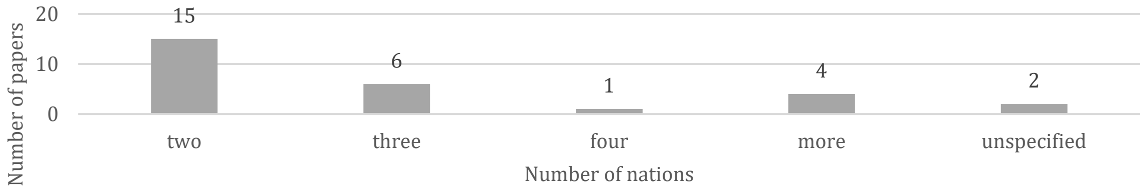

Figure 3-3 – Number of Nations Compared in the Papers

Figure 4 illustrates that next to the United States, China was the second most compared nation. In fact, the United States was often compared directly with China (e.g., Chang et al., 2014; Fang et al., 2011). In addition, the United States was also compared to South Korea four times (e.g., Dinev et al., 2009; D. J. Kim et al., 2016). The comparison between the United States and China or South Korea was usually based on the argument that the United States is very individualistic (91 on Hofstede's scale) and China or South Korea are very collectivistic (20 and 18, respectively) (Hofstede Insights, 2010). Also, a comparison between the United States and European nations occurred several times (e.g., United States vs. Italy (Dinev et al., 2006) or United States vs. France (Vance et al., 2008)). A direct comparison between a European and an Asian nation only occurred in combination with a third or fourth nation—often North American (e.g., Canada vs. Germany vs. China (Cyr, 2008) or United States vs. Germany vs. China vs. India (Hoehle et al., 2015). Throughout all of these comparisons, the authors aimed to uncover significant cultural differences, considering the distinct nature of the cultures under consideration. However, of equal interest are comparisons between two nations with very similar cultures (China vs. Japan or Germany vs. the Netherlands, e.g., by Davison et al. (2009) or Moser & Deichmann (2021), respectively). As similar countries probably also have a similar infrastructure of emerging technologies, small cultural differences can be found.

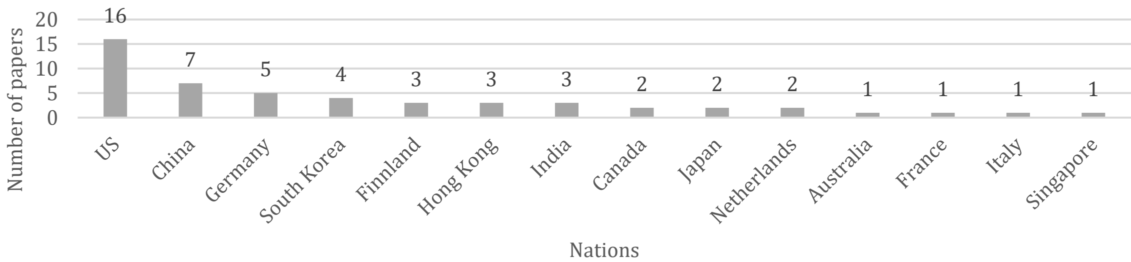

Figure 3-4 – Nations Compared in the Papers

Next, we turn to the **years of publication** and an overview of the **journals** (see Table 3). We identified that many papers on the topic of national cultural differences were published between 2006 and 2009. Subsequently, further papers were published between 2013 and 2016. Since then, the number of publications has decreased. More recent papers were found primarily as a result of the forward search beyond the basket of eight. In general, it is visible that, especially in the European Journal of Information Systems and MIS Quarterly, many papers on cultural differences have been published.

Journal/ Year of publication	2000	2001	2002-05	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019-20	2021	2022	Total
European Journal of Information Systems				1	1				1			2	3	1				1		10
Information Systems Journal					1		1				1									3
Information Systems Research																				0
Journal of the Association for Information Systems							1													1
Journal of Information Technology																				0
Journal of Management Information Systems		1				3														4
Journal of Strategic Information Systems																				0
MIS Quarterly	1			1			2							1		1				6
Other																1		1	2	4
	1	1	0	2	2	3	4	0	1	0	1	2	3	2	0	2	0	2	2	28

Table 3-3 – Matrix Journal vs. Year of Publication

Topic	Culture-related Theories										Data Collection					Data Analysis						Level		Authors
	Hofstede						Globe	Fukuyama	Hall	Others	Survey	Experiment	Interview	Delphi Study	User content Analysis	SEM (PLS separately)	SEM (interaction term)	Regression	Multi-Group Analysis	Descriptive	Other	Individual	National	
	IDV	MAS	PDI	UAI	LTO	IVR																		
Managing Organizations	x										x					x						x		(Chang et al., 2014)
	x		x								x								x			x		(Yan Chen & Zahedi, 2016)
										x	x							x					x	(Davison et al., 2009)
		x					x				x												x	(Fang et al., 2011)
	x		x		x							x										x		(George et al., 2018)
											x					x				x			x	(Gu et al., 2021)
												x										x	x	(Keil et al., 2000)
	x		x									x											x	(Keil et al., 2007)
	x				x						x							x					x	(Moser & Deichmann, 2021)
	x		x											x									x	(Schmidt et al., 2001)
										x			x							x			x	(Tuunanen & Kuo, 2015)
											x					x							x	(M. Zhang et al., 2013)
Privacy	x		x		x						x					x							x	(Cyr, 2008)
	x		x									x											x	(Cyr et al., 2009)
	x			x	x				x		x								x				x	(D. J. Kim, 2008)
	x											x							x				x	(Sia et al., 2009)
												x									x		x	(Vance et al., 2008)
															x								x	(Yang et al., 2018)
	x					x					x					x							x	(Z. Zhou et al., 2015)
Technology Acceptance	x		x		x			x			x								x				x	(Dinev et al., 2006)
	x		x		x						x								x				x	(Dinev et al., 2009)
	x		x		x	x									x	x			x				x	(Guan et al., 2022)
	x			x					x		x					x			x				x	(D. J. Kim et al., 2016)
	x												x								x		x	(Miltgen & Peyrat-Guillard, 2014)
E-Commerce	x		x		x						x						x			x		x		(Hoehle et al., 2015)
	x		x								x											x		(McCoy et al., 2007)
	x										x						x					x		(S. Sharma et al., 2022)
	x		x								x						x					x		(Srite & Karahanna, 2006)

Table 3-4 – Concept Matrix

3.5 Research Agenda

In the following, we present a research agenda guiding the way of future national culture research. We recommend extending the application scenarios, challenge culture-related theories, recommend different types of data collection and analysis, the level of measuring culture, and additional cultural research regarding more and previously barely considered nations.

Application scenarios: We identified four main literature streams concerning the effect of national culture on emerging technologies: Management of organizational processes, privacy, technology adoption, and e-commerce. But even beyond these four categories, several RQs remain unanswered. For instance, previous research identified different strategies for adopting emerging technologies and differences in managing organizational processes. Therefore, future research could investigate how to overcome national cultural differences in order to successfully implement emerging technologies (Fang et al., 2011). Especially with regard to the use of emerging technologies, such as virtual reality, a possible RQ could be how users from different national cultures accept these emerging technologies. Moreover, the VOSviewer tool has proven to be very helpful in structuring the literature and identifying the contexts. Therefore, we propose to use it in further research projects as well. In addition, dealing with ethical issues is handled differently in various national cultures. For instance, Bellman et al. (2004) found a difference in privacy concerns related to cultural values which shape a nation's regulations. Therefore, a future RQ could be how to design emerging technologies that can be used across national borders, although different data protection laws are in place and ethical preferences (HIPAA vs. GDPR) (Awad et al., 2018; Beens, 2020).

Culture-related theories: Although many papers have adopted a similar approach by conducting a survey using Hofstede's dimensions, many papers omit specific dimensions of Hofstede. Thus, some dimensions of national culture are not taken into account (Hoehle et al., 2015). For example, Chang et al. (2014) measured individualism and uncertainty avoidance between the United States and China as these differ among the culture dimensions (91 to 20 and 40 to 80, respectively). However, there is also a large difference in long-term orientation (26 to 87) and indulgence (68 to 24) (Hofstede Insights, 2010), but using several dimensions in one study leads to a higher effort, especially for study participants. Thus, researchers should consider the exclusion of each dimension carefully and support their decision with literature. In this vein, conducting studies to determine which dimensions generally affect the adoption of emerging technologies might be interesting. In addition, it is worth noting that certain authors have raised critiques regarding the application of national culture theories, specifically Hofstede's dimensions, at the individual and organizational levels. These criticisms highlight the limitation of using country-level analysis to fully explain individual level analysis (Venaik & Brewer, 2013). To overcome the shortcomings of national culture theories, we recommend the following four approaches. 1) *Individual Level Analysis*: Build the theory at an individual level, recognizing that not everyone perfectly aligns with a specific culture. Considering individual variations within cultures can provide a more nuanced perspective on cultural differences and their impact on behavior and attitudes. We suggest to involve individuals such as end-users or at least including them in the process of developing cultural models. By excluding organizational culture and focusing on end-users, a more representative view of culture can be achieved. However, it is important to remain focused on cultural influences rather than delving too much into individual personality traits. 2) *Sub-Cultural Considerations:* By incorporating individual-level measurements, it is likely that sub-cultural variations will be adequately accounted for. This approach allows for a more nuanced understanding of cultural differences within a broader national or regional context. 3) *Dynamic Nature:* It is important to recognize that cultures evolve, and incorporating this temporal aspect into the theories. 4) *Balancing Complexity:* While it is important to capture the complexities of cultures, it is crucial to strike a balance to ensure that the theories remain measurable. Finding a middle ground between comprehensiveness and practicality is necessary to maintain the applicability of the models. In addition, validating the dimensions across different countries would enhance the robustness of the framework.

Data collection: We identified that surveys are the most commonly used method to collect data. Although some studies acquire individuals or IT managers, we noticed that many studies question students. This can be problematic since students might adopt international preferences and not behave according to the national culture measured by Hofstede (D. J. Kim, 2008). This may affect the assumed influence of national culture. In addition, this can be similar in online environments where individuals might adopt other preferences than the anticipated national culture. Therefore, we briefly point out that it should be carefully considered which participants are questioned.

Furthermore, as nearly all papers found national cultural differences, it would be interesting to use complementary research methods (Davison et al., 2009), for example, by conducting qualitative methods for finding reasons and solutions for problems such as misunderstandings generated from national cultural differences. Another data collection approach might be longitudinal approaches to identify the change of culture over time.

Data analysis: As evident from the findings, three approaches based on the PLS model are commonly employed. Based on this result, we recommend a multi-group analysis to determine the differences between the national cultures. While this is a valid approach for further cultural research, we also call for using additional analyses. One recommendation for a different analysis approach would be hierarchical linear modeling to examine the influence of different cultural groups on the adoption of emerging technologies (Keil et al., 2000). In addition, Srite and Karahanna (2006) suggest that measuring the influence of the Hofstede dimensions with each other could contribute to research. As the impact of national culture has been overlooked in recent years, it would be interesting to include national culture as a moderator for the adoption of emerging technologies like artificial intelligence, virtual reality, and the metaverse since these technologies allow collaboration and working across borders.

Level: Myers and Tan (2002) have already noted that culture is not a straightforward concept. For instance, a nation might have more than one culture (e.g., India), and culture changes over time. Thus, they conclude that national culture cannot be treated as an equivalent to a nation (Myers & Tan, 2002). Thus, Srite and Karahanna (2006) suggested that the culture dimensions should not be defined at the national level using the nation as a proxy but rather measured at the individual level. This has the advantage that if individuals do not entirely identify with their given culture or are part of an ethnic subgroup, it can be captured as an influence. Hoehle et al. (2015) also highlighted the difference between those two types of measuring culture by testing both approaches and discovering the differences. Although some followed this direction (Chang et al., 2014; Yan Chen & Zahedi, 2016; George et al., 2018; Hoehle et al., 2015; S. Sharma et al., 2022), we emphasize continuing this approach in future research. In addition, the influence of groups on emerging technologies could be interesting in the context of national cultural research, thus, the individual is neither solely questioned nor is a national culture assumed (Keil et al., 2007).

Additional Research Directions: In addition, comparing more than two nations could allow a more precise assessment of national cultural differences. Keil et al. (2000) provided a useful example in this context. They included Singapore (low uncertainty avoidance index), the Netherlands, and Finland (both medium UAI, according to Hofstede). An extension involving a nation characterized by a high uncertainty avoidance index could yield intriguing findings. Another comparison could be between similar nations that differ only in one dimension. This would replicate findings (D. J. Kim, 2008) and improve the quality of measuring all dimensions by keeping similar dimensions tight and only differences between the other dimensions become apparent. Furthermore, extending to rarely considered nations, such as the parts of Asia, Arab, or African region, could also provide further exciting details. Lastly, a quantitative analysis of the effect of culture on emerging technologies could reveal overarching findings (i.e., meta-analysis).

3.6 Limitations and Contributions

National culture is an important influencing factor on emerging technologies in managing organizations, privacy, technology adoption, and e-commerce. We identified several characteristics that are typical for the measurement of culture in IS literature which are presented in a concept matrix in Table 4. For example, the majority of the papers applied Hofstede's Cultural Dimensions Theory and used surveys in a similar way. In contrast, there are differences in the measurement of national culture, especially on the analysis level. Therefore, we attempt to decipher ways to measure the impact of national culture on emerging technologies. Finally, we present the results in form of a research agenda to guide future research to further topics, extending existing theories, and testing new data analysis methods.

Although we have taken great care in reviewing the literature, this paper is subject to some **limitations**. Since we examined the RQ qualitatively, our study is subject to certain restrictions. First, the literature review results are significantly influenced by the keywords chosen. We have taken meticulous care in selecting a comprehensive set of keywords that encompasses a wide range of potential papers. However, it is important to note that alternative keywords might yield diverse outcomes. Second, our structured literature review was primarily based on the top IS journals, commonly referred to as the "basket of eight". Nonetheless, it is worth considering the application of these keywords to other databases and across various disciplines in future research. Third, our particular focus was on measuring the impact of national culture as a moderator. However, it would be beneficial for future studies to broaden this scope and explore additional dimensions, such as investigating the measurement of national culture as an independent variable. Finally, although we used the VOSviewer tool to develop the categories in "application scenarios" (see Figure 2) to minimize the researcher's subjective influence in the selection, the identified categories of the concept matrix can, of course, be extended with other elements or categories in further studies.

Our **theoretical contribution** to IS research is providing an overarching IS-centric theory of how national culture influences the adoption of emerging technologies conceptualized by the concept matrix. This matrix offers comprehensive insights into the common utilization of culture-related IS theories, as well as the diverse research methods employed. Second, it also helps the IS research community to identify research gaps. Thus, we developed a research agenda to pave the way for further research. In this vein, we organize the previous literature and make implicit knowledge explicit. In particular, we highlighted potential research streams in regard to the effect of national culture on emerging technologies, expanding the theories, and finally emphasized the community to use novel empirical methodologies (e.g., hierarchical linear modeling). In particular, we call for examining national cultural differences and similarities between cultures on an individual level since collaborations are becoming more important due to globalization and more opportunities to work across borders (e.g., emerging technologies such as metaverse and AI). From a **practical perspective**, we first contribute to highlight the importance of national cultural understanding. Organizations operating across national borders and serving customers from diverse cultural backgrounds need to comprehend national cultural differences to effectively meet their needs. Recognizing and comprehending national cultural variations is crucial for fostering teamwork, successfully adopting emerging technologies, and retaining and attracting international customers. Organizations benefit from the presented knowledge to perform their own analyses with respect to introducing emerging technologies, i.e., new sales markets in other cultures or nations. Second, since it is difficult to manage globalized organizations, we contribute to practice by presenting that difficulties can arise when multinational organizations ignore national cultural differences and try to apply management formulas developed in their own cultural context. Thus, there is no one-size-fits-all approach for running an organization and adopting emerging technologies, considering the impact of cultural differences. Third, the organizations can explore the collected literature for relevant application scenarios to identify what cultural influences have already been observed and what recommendations are made by the respective authors. Thus, they can tailor their plans and strategies according to the effect of national culture to maximize their competitive advantage.

3.7 Acknowledgements

This research and development project is/was funded by the German Federal Ministry of Education and Research (BMBF) within the "Innovations for Tomorrow's Production, Services, and Work" Program (funding number 02L19C150) and implemented by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication.

4 Paper B: An Industry-Specific Investigation on Artificial Intelligence Adoption: The Cases of Financial Services and Manufacturing

Title

An Industry-Specific Investigation on Artificial Intelligence Adoption: The Cases of Financial Services and Manufacturing

Authors

Hoffmann, Mathis and Mehler, Maren F.

Publication Outlet

Proceedings of the Pacific Asia Conference on Information Systems (PACIS 2023), Nanchang, China

Abstract

Artificial Intelligence (AI) has a lasting transformational effect on industries worldwide. Former research has primarily focused on AI adoption as a business phenomenon without considering different industries. Those are characterized by unique attributes that may influence how modern technologies are implemented. In order to initiate non-generalized research in that field, industry-specific drivers and barriers to firm-level AI adoption in the financial services and the manufacturing industry are analyzed. Drawing on the Technology-Organization-Environment (TOE) framework, it was possible to paint a holistic picture of use cases and unique, but also general drivers and barriers of AI adoption for each industry. Ultimately, by bringing these two viewpoints together, a theory of hard (generalizable) and soft (industry-specific) AI adoption factors was developed. Therefore, the findings serve as a basis for further industry-specific research and provide business stakeholders and executives with a transparent handbook about industry insights and AI knowledge.

Keywords

Artificial Intelligence, Adoption, TOE Framework, Financial Services Industry, Manufacturing Industry

4.1 Introduction

Most companies know that Artificial Intelligence (AI) offers great potential for saving time and costs, thus increasing profits. For example, 34% of companies worldwide are already using AI in 2022, and 42% are exploring AI (IBM, 2022b). AI is estimated to contribute $15.7 trillion to the global gross domestic product (GDP) in 2030 (PwC, 2017). AI embraces several disciplines and techniques in computer science, but business leaders' focus has also shifted to how the technology's capabilities can be leveraged to create economic value. Despite the many opportunities AI offers, many companies are—if at all—still exploring and have not yet adopted AI. Two reasons can explain this: First, barriers, if they exist, hinder the implementation, and second, potentially missing drivers will not support the adoption process of AI in organizations (Pumplun et al., 2019; Zöll et al., 2022).

Research on organizational and individual technology adoption is not new. The most established frameworks are the Technology Acceptance Model (TAM) (Davis, 1989) and Technology-Organization-Environment (TOE) (DePietro et al., 1990). In the last years, the investigation of the so-called 'AI readiness factors' has also become established, i.e., the organizational 'chassis' embracing all prerequisites for enablement of successful AI adoption. Respective contributions have been performed by Jöhnk et al. (2021) and Pumplun et al. (2019) with qualitative, interview-based methods. Readiness factors can be seen as a preceding level before the organizational adoption of AI itself. However, the results were aggregated and generalized across all industries.

Fundamental strategic management science, like Porter's five forces (Porter, 1980), argues that a company's industry structure is characterized by different, maybe even unique industry parameters. This is why technology adoption factors might also be affected and differ between industries, opening up the need to zoom in on selected industries. This need has also been recognized by authors of previous studies (Cubric, 2020; Kar et al., 2021; Pumplun et al., 2019; Zöll et al., 2022). However, industry-specific research on drivers and barriers to AI adoption only exists in a few cases, and a cross-industry comparison is missing. Kruse et al. (2019), for example, answered the question about the challenges and barriers of AI in banking and insurance by conducting interviews with AI experts from the German finance industry and supporting industries. Kinkel et al. (2021) defined a questionnaire-based survey and collected responses from over 600 industry stakeholders to investigate factors of AI adoption in the manufacturing industry.

Thus, this paper seeks to contribute to the research on AI adoption in financial services and manufacturing from separate viewpoints. The financial services industry offers great potential (Hentzen et al., 2021). Manufacturing is very diverse and constantly in transformation and, therefore, has many use cases (de Propris & Bailey, 2020; United States Department of Labor, n.d.; X. Zhang et al., 2019). The holistic perspective on drivers and barriers in both industries intends to exert an industry-specific reality check of generalized, industry-aggregated research. Therefore, we seek to answer the following research questions (RQs):

RQ1: What factors drive AI adoption in the financial services and manufacturing industries?

RQ2: What factors curb AI adoption in the financial services and manufacturing industries?

RQ3: How do drivers and barriers of AI adoption relate to each other across industries?

To answer the research questions, we conducted an interview-based multiple-case study. The process of qualitative data collection aims to establish verified causal findings and explanations for the described phenomena. The TOE framework, a widespread theory in technology adoption research, supports the research. With this paper, we urge future research on the adoption of technologies to consider industry-specific factors by showing that the financial service industry and manufacturing industry have different drivers and barriers and vary from known general factors.

This paper is structured as follows: After the introduction with our motivation and research questions, we present the TOE framework, fundamentals of AI, and related work. We then describe our case study methodology in chapter three and provide the results of our research according to the TOE framework in the next chapter. Finally, we discuss the results, give theoretical and practical contributions and limitations, and conclude our paper.

4.2 Theoretical Background

4.2.1 Technology-Organization-Environment (TOE) Framework

Technology adoption is not an isolated phenomenon but represents "a series of decisions that are not visible to all [involved] participants" (DePietro et al., 1990, p. 178). To ensure that all influencing factors are taken into account, it is established in research to use technology adoption frameworks, which can support a complexity-reducing abstraction. The TOE framework is one of the most common technology adoption frameworks in information systems (IS) research (Kinkel et al., 2021) focusing on firm-level and including firm-external influences (Oliveira & Martins, 2011). It was utilized, for example, for adoption technologies such as Big Data (Bremser, 2018), of AI (Pumplun et al., 2019), but also for specific AI adoption barriers in finance (Kruse et al., 2019). The TOE framework identifies three different contexts in and surrounding a company that influence technology adoption: the technological context, the organizational context, and the environmental context (DePietro et al., 1990) (see Figure 1).

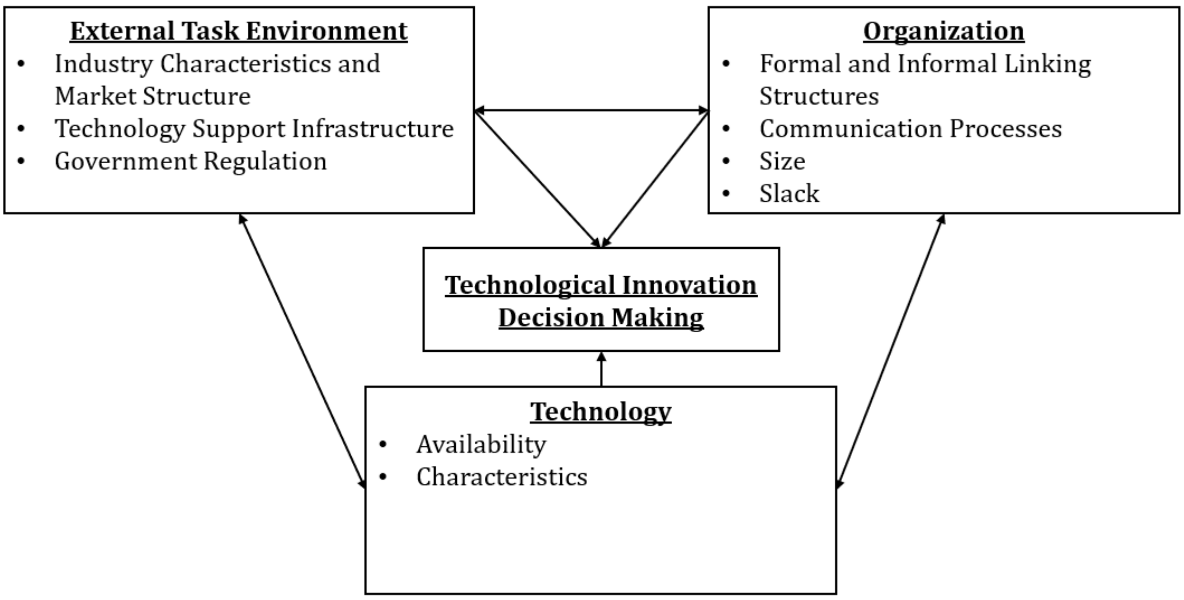

Figure 4-1 – Technology-Organization-Environment (TOE) Framework, According to DePietro et al. (1990)

The **technological context** embraces internal and external technologies that are or may be relevant to a firm. The focus is on how an innovative technology itself can influence a company's adoption process. This includes AI and its technical requirements, respective sub-disciplines, but also alternative technologies, and the holistic enterprise IT infrastructure. The **organizational**

context covers a firm's descriptive measures: e.g., firm size, managerial structure, or human resources. The **environmental context** is the 'arena' in which a company operates, bringing opportunities and constraints for technological innovation (DePietro et al., 1990). The TOE framework has a solid theoretical basis and consistent empirical support in IS research (Oliveira & Martins, 2011). In AI adoption research, the TOE framework was already applied multiple times: Kruse et al. (2019), for example, based their analysis of AI adoption challenges in the financial services industry on the TOE framework. Kinkel et al. (2021) even explicitly concluded that the framework was suitable for AI. Other applications can be found, for example, in the area of AI readiness factors (Jöhnk et al., 2021; Pumplun et al., 2019). Therefore, the TOE framework is expected to be useful also for this study in arranging the findings, interpretations and disclosing connections.

4.2.2 Artificial Intelligence

Artificial Intelligence (AI) is seen as a major innovative technology for businesses in almost every industry and will likely force companies to transform their core processes, value proposition, and business models (Brynjolfsson & McAfee, 2017; Sjödin et al., 2021). We follow Russell and Norvig (2021), describing AI as "intelligent agents that receive percepts from the environment and take actions that affect that environment" (Russell & Norvig, 2021, p. 7). They also define Machine Learning (ML) as a subcategory of AI which gives predictions based on experience. While companies mainly adopt ML, other types of AI should not be neglected here. Therefore, the term AI will be used in the following in order to remain more generic.

To understand AI adoption as a transformational journey, drivers and barriers can be distinguished. Former research has already investigated these adoption factors but has left some gaps open: first, most research did not include an industry-specific focus (e.g., Pumplun et al., 2019), leading to generalized findings. Second, the little existing industry-specific research has not produced a holistic view of both drivers and barriers (e.g., Kinkel et al., 2021; Kruse et al., 2019). Third, specific ways of looking at at least two industries have not yet been brought together to reveal relations, interdependencies, and commonalities. This paper will analyze the financial service and manufacturing industries, as many AI use cases can be found (Hentzen et al., 2021; X. Zhang et al., 2019). Also, different use cases exist: While the financial services industry has a focus on customers and intangible products and services, the manufacturing industry focuses on physical products. Thus, different drivers and barriers are expected.

Financial services companies are subject to special regulatory requirements (Freudenstein et al., 2019; Kapsis, 2020). In addition, the IT infrastructure is dominated by technologies from the last decades, so-called 'legacy IT' systems (Freudenstein et al., 2019). Furthermore, the industry is characterized by strong competition among other FinTech start-ups offering personalized services and achieving lower costs through AI (Kruse et al., 2019; I. Lee & Shin, 2018; Sinn et al., 2021). The potential of AI ranges from the back-end, noncustomer-facing operations to front-end, customer-facing scenarios, e.g., detecting consumer credit delinquencies (Khandani et al., 2010); assess customers' credit-worthiness (Dastile et al., 2020); chatbots to automate customer interaction (Riikkinen et al., 2018) or as 'robo-advisors' for automated portfolio and investment management (Buchanan & Wright, 2021; Hentzen et al., 2021).

Influencing effects on the adoption at the firm level have barely been analyzed in the industry alone and especially not as a comparison between industries. Kruse et al. (2019) assessed the challenges that financial services companies face when planning to adopt and implement AI.

Strong similarities to generic barriers to AI adoption can be found, but challenges in the financial services industry seem more diverse (see Table 1).

The manufacturing industry is in constant transformation, as 'industrial AI'—mapping AI application scenarios onto manufacturing use cases (X. Zhang et al., 2019) —enables process optimization (Peres et al., 2020), quality control (Ding et al., 2020; Ojer et al., 2020), predictive maintenance and human-robot collaboration (Peres et al., 2020). Thus, just like in financial services, the question arises of how AI adoption is accelerated and hindered in the industry. Kinkel et al. (2021) analyzed manufacturing companies' use of AI and TOE-classified adoption variables through a survey among more than 600 participants from the manufacturing industry.

Table 1 shows generic drivers (D) and barriers (B) according to Kar et al. (2021), specific barriers for the finance industry according to Kruse et al. (2019), and specific factors in the manufacturing industry according to Kinkel et al. (2021) (G—General, FS—financial services and M—manufacturing) that former research comprised. Additionally, these factors were clustered according to the TOE framework (T—Technology, O—Organization, and E—Environment).

Variable	Description	Industry	Factor	TOE
Accuracy	Improved accuracy in decision-making and forecasting.	G	D	O
Cost-reduction	Automated business processes, efficient decision-making, and reduction of human error.	G	D	O
Decision-making	Automated, data-driven decision-making.	G	D	O
Productivity	Improved productivity and efficiency in business processes.	G	D	O
Speed	Improved time-to-decision due to automation.	G	D	O
Well-being	Less workload and reduced stress for employees.	G	D	O
Sustainability	Sustainable processes and supply chains.	G	D	O, E
Customer experience	Service improvement by learning customer preferences.	G	D	E
Data	Lack of data quality and data quantity.	G, FS	B	T
Infrastructure	Infrastructure support is required for wide-scale implementation. Legacy IT needs to get replaced.	G	B	T
Model reusability	Reuse of the AI model for different problem scenarios is difficult.	G	B	T
AI strategy	Lack of an AI strategy defining how AI should be used to meet business goals.	G	B	O
Job security	Streamlining of routine-based job profiles.	G	B	O
Leadership	Lack of leadership commitment towards AI adoption.	G	B	O
Trust	Lack of trust in AI-generated decisions.	G	B	O
Use case identification	In some cases, AI solutions are less effective than traditional ones. Return on investment is at risk.	G	B	O
Know-how	Labor market gap or lack of budget to attract AI talent.	G	B	O, E

Variable	Description	Industry	Factor	TOE
IT infrastructure	Legacy IT needs to get replaced.	FS	B	T
Vague market	Low availability of suitable AI software.	FS	B	T
AI characteristics	Risk, compliance, ethical standards.	FS	B	T, O
Agility to adapt	Adapt firm resources (financial, technical, human).	FS	B	O
Changing process competencies	Change management, changing environment.	FS	B	O
Know-how	Professional expertise regarding AI skills.	FS	B	O
Organizational hierarchy	Lack of organizational agility.	FS	B	O
Top management support	Lack of support by decision-makers.	FS	B	O
Competitive pressures	Especially driven by FinTechs.	FS	B	E
Data protection	Data protection is part of the core business.	FS	B	E
Governmental regulation	Regulatory requirements (BCBS239, MiFID).	FS	B	E
Lack of customer support	Moral concerns by customers.	FS	B	E
Company size	Measured by the number of employees.	M	D	O
Design	Role of product design as a competitive strategy.	M	D	O
Digital skills	Corporate technological know-how.	M	D	O
Product quality	Role of product quality as a competitive strategy.	M	B	O
TOE elements: T - Technology, O - Organization, E - Environment; Factor: D - Driver, B - Barrier; Industry: G - General, FS - Financial Services, M - Manufacturing				

Table 4-1 – Drivers and Barriers of AI Adoption From Literature

Overall, drivers and barriers of adoption variables exist (see Kar et al., 2021; Kinkel et al., 2021; Kruse et al., 2019), but their focus is on general AI adoption instead of individual industries and a cross-industry comparison. The next chapter describes how the qualitative research design of this study intends to target this gap.

4.3 Methodology

As the research questions require the construction of an exploratory and explanatory model, interview-based case studies provide the best opportunity to analyze phenomena regarding their causality and inherent interrelationships. The structure is oriented on the combination of the five case study design components of Yin (2014) and the step-by-step chronological framework by Eisenhardt (1989).

Getting started. A case study should have clear, predefined research questions to be used for theory-building (K. M. Eisenhardt, 1989), serving as the first out of five central case study design components (Yin, 2014). Therefore, the research questions were developed iteratively by reviewing existing literature in scientific databases and a forward and backward search. The state of research on both generalized and industry-specific AI adoption was derived. These account for the so-called 'theoretical study propositions', the second design component (Yin, 2014).

Selecting cases. A multiple-case study offers analytical generalization of the findings by deriving substantiated theories between cases (Yin, 2014). This results in external validity (Yin, 2014). While a case study cannot reflect variation in every possible dimension (Yin, 2014), some degree of environmental and characteristic variation should be considered while planning the investigated cases instead of applying random selection (K. M. Eisenhardt, 1989). Thus, the definition of the study's unit of analysis (i.e., the 'case') was set up as the third research design principle (Yin, 2014). In this paper, a case is defined as an individual with expertise in topics related to organizational AI adoption, working for either a private company within a focus industry (manufacturing and financial services) or holding profound stakes and insights about it. This strategy makes use of potentially different perspectives on the topic and supports the idea of replication logic. A detailed list of cases in the data collection is presented in Table 2. This table includes information about the current role, overall job experience, experience in current role, location, and company size. The interviews were conducted in July and August 2022; the average duration was 27 minutes (min 24, max 32 minutes), and written answers were given for one interview.

#	Industry	Role	Experience (in current role)	Location	Size
Financial Services					
I.1	Management consulting	Managing consultant—AI in financial services	10 (4) years	UK	L
I.2	Insurance	Head of AI solutions	15 (5) years	Germany	M
I.3	Insurance	Head of AI and data development	24 (1.5) years	Germany	S
Manufacturing					
I.4	Management consulting	Senior consultant—AI in manufacturing	7 (5) years	UK	L
I.5	Management consulting	Senior consultant—digital operations	4 (Unknown)	UK	L
I.6	Management consulting	Managing consultant—intelligent automation and AI	13 (2.5) years	Germany	S
I.7	Mechanical and plant engineering	Product manager—pre-sales digital solutions	8 (3) years	Germany	M
Size: S - < 5,000 employees, M - 5,000 - 50,000 employees, L - >50,000 employees					

Table 4-2 – Interview Partners

Crafting instruments and protocols. The next steps are a clear definition of the sources for data collection, the types of collected data, and the number of investigators performing research (K. M. Eisenhardt, 1989). This paper's data collection only included qualitative data without any additional case-related documents. One author performed the collection.

Entering the field. All interviewees received the questionnaire in advance to ensure a smooth interview. In case study research, data collection and data analysis are not strictly separated (K. M. Eisenhardt, 1989). Thus, writing down field notes and impressions during the investigation is required to enable adjustments to the interview question protocol and provide flexibility to respond to emerging trends (K. M. Eisenhardt, 1989). These adjustments in interview questions were essential in the early stages of data collection, as some optional questions turned out to be obsolete. In contrast, additional questions in the following interviews could deepen other emerging and surprising topics.

Analyzing the data. The analysis was prepared by defining an analytical strategy and choosing analytical techniques. The setting of these parameters forms the fourth design component. An analytical strategy is important for the fair treatment of evidence and for producing logical analytical conclusions (Yin, 2014). Out of four common strategies in case study science (Yin, 2014), three can be ruled out based on the scope of this study. The chosen strategy lets the theoretical propositions lead the study and is the most common one in research (Yin, 2014). This strategy uses the theoretical foundations as external sources of knowledge to relate the findings of this study. This makes it possible to fit new data into the findings from the literature. Conceptually, the pattern-matching technique has been applied as a guide for qualitative analyses. This technique relies on using codes, i.e., categories representing a piece of thematic content. First, the interview transcripts were divided into two groups according to the industry. Consequently, the following steps were performed for each group separately. Second, several deductive codes (main categories) were brought to the interview data, yielding a content-related structure and logical parts. This was followed by deriving inductive codes, born out of the specific content of the interviews, and breaking the main categories down into subcategories. These subcategories enabled mapping sets of cross-case interview statements onto sole findings within an industry group. However, starting this process again for each case resulted in new subcategories for each interview transcript. To ensure that no later-defined subcategory was missed in already-reviewed transcripts, steps 3-5 were repeated iteratively until the number of codes converged (Kuckartz & Rädiker, 2022).

Shaping hypotheses. The explanation-building technique was applied to summarize and report the findings (Yin, 2014). Therefore, an iterative process was applied between adapting and comparing a statement with the qualitative data (Yin, 2014). The following order facilitated this: A list with all text segments within the group carrying the focal code was generated for each code. With this code-relevant overview, it was possible to analyze derivable key messages iteratively and identify to what extent summarizing was possible and where separation was needed.

Closing research. Research is considered complete when saturation of results is achieved and incremental growth is minimal. Although only three (respectively four) interviews were conducted within the industries, hardly any new insights resulted from the last interviews (in both industries, only one, respectively two new inductive codes could be assigned in the analysis of the last interview in each case, and it became increasingly difficult to separate the content from existing codes), thus, saturation was achieved.

Enfolding literature. Identifying and dealing with literature conflicting with one's findings—so-called 'rival explanations'—is a central step in establishing internal validity and assessing its generalizability (K. M. Eisenhardt, 1989), yielding in the implementation of the fifth and last research design principle (Yin, 2014). Generalizability needs to be understood as the question of whether the study findings are scientifically complete. This was tested by using a two-sided strategy. First, an outside-in technique was used to identify previous research findings in this study's results. Afterward, an inside-out technique was used to identify this study's results in previous research findings. As a result, gaps and differences were disclosed and enabled a basis for critical reflection.

Besides the five fundamental design components, case study research should comply with four general quality tests (Yin, 2014). Each test can be complied with by realizing a set of techniques and tactics inside the research design (Yin, 2014). Table 3 gives a respective overview together with a summary of their implementation in this study.

Test	Theoretical case study tactic	Tactic implementation in this paper
Construct validity	Use multiple sources of evidence.	Cases included internal and external industry stakeholders to use different perspectives.
	Establish chain of evidence.	Detailed, chronological description of the method enables to trace back all steps and conclusions.
	Have key informants review the draft case study report.	Transcript review was optionally offered for review to the interviewees, but the offer has not been used.
Internal validity	Pattern matching.	Iterative text coding.
	Explanation building.	Iterative shaping of conclusions out of the data.
	Address rival explanations.	Addressing former research and identifying gaps.
	Use logic models.	Not applicable.
External validity	Use replication logic in multiple-case studies.	Iteratively shaping conclusions out of the data.
Reliability	Use case study report.	See Results chapter.
	Develop case study database.	Composition of case overview, interview transcripts and case study report.

Table 4-3 – Case Study Quality Tests According to Yin (2014)

4.4 Results

This chapter is divided into four parts. First, the interviewees' perspectives on AI adoption and use cases are assessed for both the financial and manufacturing industries. In the second and third sections, the drivers and barriers of each industry are presented and categorized according to the TOE framework. In the fourth section, the drivers and barriers of both industries are compared.

4.4.1 State of AI Adoption and Use Cases

4.4.1.1 Financial Services Industry

Financial services companies seem to be aware of AI's business potential, as I.2 and I.3 both are the heads of official AI-specialized organizations within their company. Also, I.1 believes that most companies at least state to target AI capabilities in their corporate strategy. However, the actual state of implementation seems to be in its infancy. Many financial services companies are experimenting with AI in so-called 'sandbox environments' (I.1), that is, in isolated systems where software can be tested safely. Only one respondent mentions that his company is leading in AI adoption compared to others he is in touch with through several cross-company AI committees (I.2).

AI use cases can be classified into two categories: 'Offensive' use cases transform the way business is run, e.g., how market interaction looks like and which services are being offered (I.1). In contrast, 'defensive' use cases target improving existing processes and due diligence activities and making them more secure (I.1). Overall, many use cases focus on customer analytics and customer behavior, such as risk management, or business operations, such as algorithmic trading (I.1; I.2; I.3).

4.4.1.2 Manufacturing Industry

The manufacturing industry structure is more diverse compared to financial services. As a result, the identified use cases and process applications are wide-ranging. Some interview partners state that the use of AI is in its infancy and is treated cautiously—including large, established companies that are still completely at the beginning (I.5). Here, AI is applied only in selected individual cases, which may even be separate from one another (I.6). Others have witnessed a wide range of AI in various use cases (I.4; I.6). The most common scenario is the field of quality control by combining hardware, software, and intelligent algorithms to automate the manual and labor-intensive inspection of manufactured parts (I.4). Besides that, the respondents presented applications of predictive maintenance or scheduling tasks on shop floors (I.4; I.7). However, a holistic approach of extracting, processing, analyzing, and using data along the entire value chain is hardly found in any manufacturing company (I.6).

4.4.2 Drivers of AI Adoption

Next, the drivers of AI adoption derived from the interviews shown in Table 4 are presented. Eleven drivers are found for the financial services industry and nine for the manufacturing industry. In the following two sections, these factors are categorized and explained according to the TOE framework.

4.4.2.1 Financial Services Industry

Technological drivers. As financial services providers are confronted with extensive demands on their data management, these regulatory requirements do naturally exist as part of their business models and significantly increased after the global financial crisis of 2008/2009 (I.1). However, they transparently highlight possibilities of automation and other uses of AI, making its application easier (I.1). AI adoption is also driven by standardizing processes for scalable, reusable, and flexible AI systems (I.3). That means building a technological baseline in the form of a system that can be used again in another context. Hand-in-hand with this effect is the opportunity of realizing leverage effects (I.3), which occur when the implementation of a single use case offers the ability to have a disproportionately bigger advantage to the organization compared to the development effort.

Organizational drivers. Almost all organizational drivers can be summarized as business decisions aimed at either growth or cost reduction. According to the interview partners, the organizational intention to reduce costs can primarily be found in defensive use cases. This can be achieved by optimizing, automating, and streamlining existing processes (I.1; I.2; I.3), for example, by using intelligent recommendation systems to minimize search times in internal systems for clerks (I.2). A company's effort to maximize profits, in particular, is reflected in the offensive applications of AI (e.g., customer analytics, recommendations and cross-selling, algorithmic trading). Moreover, AI adoption is accelerated by the trend toward data-driven decisions. Forecasting models, in particular, can serve as decision support systems in order to help executives back up their decisions with numbers and facts rather than relying on belief and gut feeling (I.3).

Environmental drivers. Compared to the organizational context, environmental drivers are far more diverse, as they originate from several actors within the market. First, the use of AI is accelerated by the desire to improve customer experience (CX) (I.2). Again, a kinship with defensive use cases can be noticed, as these can be leveraged to optimize CX journeys by improving or automating processes and reducing execution times. Besides the customers, the market environment and competition could be identified as having accelerating effects on AI adoption. However, this aspect must be considered from two perspectives: On the one hand, the respondents referenced competitive pressure within the market (I.2; I.3). On the other hand, cross-company exchange (e.g., through committees and conferences) allows theoretical ideas to bear fruit (I.2; I.3). Surprisingly, also regulatory requirements imposed by the legislator have an accelerating effect on AI adoption. These requirements can be so overwhelming that companies hope to be able to handle them by using AI (I.1). "For example, there is a large bank […] [facing the problem] of money laundering. They are hoping to get a better handle on their money laundering problem by using AI because they realize that it's just too gigantic. These are too gigantic of an undertaking for humans to do." (I.1)

4.4.2.2 Manufacturing Industry

Technological drivers. Not a single technological driver for AI adoption could be identified in the manufacturing industry.

Organizational drivers. Implementing AI started with curiosity a few years ago. In fact, curiosity towards AI and its hype as a revolutionary technology with value-adding potential was and still is one main driver for AI adoption (I.5; I.7). This is intertwined with the sheer will to become digital leaders, resulting in a high degree of intrinsic commitment to AI (I.7). One of the effects that manufacturing companies hope to achieve when applying AI is increasing productivity (I.4). This is accompanied by other factors that have a similar impact on business operations. These include efforts to reduce or automate labor-intensive, repetitive, manual, complex, and expensive operations (I.4; I.6). For example, the application of predictive maintenance falls into this category, as downtimes, and thus costs can be reduced (I.7). In addition, manufacturers also strive for consistent quality of the products produced (I.6), which again can also be simplified or improved with the help of AI—as described earlier in the quality inspection use case. "Maybe another driver here is to look at 'what are the low-hanging fruits?' And I will assume to some extent that this is one of the reasons why AI is used so often in quality control. You are not actually designing the process. All you are doing is adding an intermediate station, so to speak, that takes a quick look with a camera to see if everything looks okay. That does not affect the business and can only be a win." (I.6) In conclusion, organizational drivers stem almost entirely from economic intentions. As a rule, a business case serves as the basis, which in the case of a positive return on investment (ROI) is the clear leader among the AI adoption drivers (I.7).

Environmental drivers. Surprisingly, competitive pressure and the goal of achieving competitive advantages seem to play a minor role when it comes to AI adoption in manufacturing (I.5). However, the company's ambition to become part of a circular economy (CE) is crucial for the transformation to a data-driven company with a holistic AI approach (I.6).

4.4.3 Barriers of AI Adoption

In this section, the barriers to AI adoption derived from the interviews shown in Table 4 are presented. 13 barriers are found for the financial services industry and 17 for the manufacturing industry.

4.4.3.1 Financial Services Industry

Technological barriers. Adopting AI and implementing use cases starts with creating the technological conditions in the first place. In the financial services industry, I.1, I.2, and I.3 all mentioned being confronted with problems related to legacy IT as being highly self-contained and written in ancient programming languages. These systems are not suitable for integration with modern AI applications based on programming languages like Python, as they do not meet the novel idea of microservices architecture and therefore are a blocker for free and dynamic data flows (I.2). Also, cloud-based technologies, which can facilitate AI deployment, have not gotten through to financial services companies either (I.1). This, but also the cultural influence (see organizational barriers), might be the reason why the interviewees experienced a lack of tools to develop intelligent applications as another barrier of adopting AI (I.1; I.3). Data is also the central requirement for training AI. One problem is data processing into a desirable state regarding their form and quality (I.1). In addition, AI projects are hard to plan due to the unpredictability of training processes. (I.2). This makes it more difficult to forecast project milestones and estimate governance, approval, and coordination processes (I.2). I.2 argued that the technical implementation itself does also hold its challenges, but still is more plannable and forecastable than project timelines. Moreover, after implementation, it is still possible that an AI model does not live up to its expectations due to a lack of accuracy (I.2). What also hinders financial services companies from implementing highly autonomous intelligent decision-making systems is the issue of explainability (I.1; I.2). An AI system is defined to be 'explainable' when it can explain the rationale behind its decision (Samek et al., 2017). It is important for humans to verify the system, improve it, and learn from it, but also comply with legislation regarding the exact context of the application (Samek et al., 2017) instead of being treated as a black box algorithm. However, implementing explainability can be complex (I.2).

Organizational barriers. In the organizational context, the status quo of a risk-averse corporate culture is not only a barrier itself but also a root for several other challenges (I.1). "So when you are working with a new technology like […] [AI], then maybe you also have to try things out a bit faster, i.e., fail-fast, also fault tolerance. Banks and insurance companies work with data, figures, and risks every day. Rough generalization, but they tend to be more risk-averse than many other companies" (I.1). This might be a reason why outdated but reliable legacy IT is hardly getting replaced and using open-source software standards for AI development is a hurdle. Also, the collision of work types is a barrier, as data scientists and IT professionals tend to move in an agile environment, which is unfamiliar to many executives and business teams (I.1). The corporate culture may also cause a barrier to AI adoption on an even higher level, as I.1, I.2 and I.3 have experienced a lack of leadership commitment toward AI. In general, the organizational approach to the topic and identifying the strategic value of AI are key to successful adoption, which necessitates the need for an AI strategy to find company-specific answers to a variety of questions (I.1). Smaller companies, in particular, tend to struggle with answering these questions (I.2) and, above all, with answering the question of the sense of using AI at all (I.1). Finally, there are also challenges regarding the field of corporate know-how (I.1). Especially in banks, there is a trend of outsourcing most AI projects, which may be a strategic maldevelopment, as it will keep data science and AI know-how excluded from the firm (I.1).

Environmental barriers. In financial services, the biggest hurdle regarding the implementation and use of AI software consists of government-imposed regulatory requirements and compliance standards (I.1). EU-based banks and insurance providers face additional complexities regarding their data management in order to comply with European General Data Privacy Regulation

(GDPR) and to make data protection processes efficient (I.2; I.3). For this reason, financial institutions have set up entire units within the company to deal with data management (I.1).

4.4.3.2 Manufacturing Industry

Technological barriers. A special issue for manufacturing companies is the development of AI applications near physical infrastructure, such as machines or robots in the real world. This also has organizational implications, as intra- or even inter-organizational collaboration is required (I.7). Another issue is the procurement of complete smart systems. In particular, such systems apparently would be either very expensive or very inflexible (I.4). Moreover, in order to make AI models work in the first place, manufacturing companies face the issue of data availability (I.7). A lack would result in low accuracy in the model and might even make the application of the model completely useless for the specific use case (I.4). Moreover, AI may not even be the best solution for certain problems. This follows the plea of I.6 for thoughtful use of AI. Separately from this, one of the interviewees also feels that the revolutionary speed of technological developments can be too much for companies to cope with (I.4). Beyond AI algorithms, this applies above all to complementary technologies, such as hardware (I.4). Lastly, the development of AI through the use of cloud services—especially through infrastructure-as-a-service solutions—"has become so easy thanks to the existing tools. As a result, I no longer believe that this will be a major hurdle in the future" (I.6). However, the realization via the cloud becomes more difficult when robots generate high-resolution data (e.g., one data frame at a rate of four to eight milliseconds), yielding such large volumes of data that transferring them to the cloud and processing them there is hardly feasible in these cases (I.7). I.7 has an even more drastic view on the issue of cloud governance and user security concerns: The lion's share of his customers (mainly automotive companies and suppliers) do not want to deploy via the cloud and therefore are rather looking for on-premise and on-edge solutions. In conclusion, deployment seems to remain a challenge for AI adoption in the manufacturing industry.

Organizational barriers. The cautious approach to AI models running outside of the own organization (i.e., on cloud servers) likely stems from data sovereignty and internal privacy policies in manufacturing companies (I.7). In particular, a lack of standards can influence the internal development of a wide variety of solutions without transparent communication channels, resulting in a heterogeneous landscape of different AI solutions (I.6). There are also challenges at the management level. The statements here are mixed: I.4 emphasized the digitization efforts of international companies in particular, while I.7 experienced skepticism towards AI, especially among elderly and experienced managers. The unpredictability mentioned above of the success and accuracy of an AI model in the technological context has even significantly greater effects in the organizational context. These deployment imponderables are perceived as a risk and disruption to the operational process, as—in the worst case—a faulty AI system results in physical damage (e.g., "defective/damaged/useless" (I.5) goods) and thus in higher costs than in intangible application areas (I.5). In addition to this, a certain lack of failure culture is also the reason why manufacturing companies seem to be not more advanced in the field of AI (I.4). This means that executives tend to be reluctant to make an investment in the development of an AI system if the success of this system cannot be assessed from the outset (I.6). Moreover, non-acceptance or even fear from the workforce whose activities are the target of automation or workload reduction arises (I.7). This fear may be justified in some cases, but may also be due to a misunderstanding of AI itself in other cases (I.6). This concerns extreme cases in which there is no understanding of AI as a technology at all, but people rather have an undifferentiated conception (similar to science fiction) of what AI is capable of today (I.6). On the other hand, sometimes the wrong problems are solved with AI because decision-makers are unaware of which method is best suited for what (I.6; I.7). The latter again concerns the previously mentioned case of over-engineering. In conclusion, the lack of correct understanding, together with the need to build up respective know-how in the form of a data science workforce (I.7) that is merged into cross-functional, transparently communicating teams (I.4; I.7), can be an additional hurdle to adopt AI in the first place.

Environmental barriers. External regulation also seems to play only a small role for manufacturing companies. Still, regulatory requirements do affect AI adoption, but mainly in the areas of personal and customer data privacy policies—again due to the European GDPR (I.4; I.6). World affairs such as the war in Ukraine and the Covid-19 pandemic with their effects regarding sanctions on gas supply, materials shortages, drops in demand, worker unavailability, and reduced IT budgets seem to play a larger role in the environmental context (I.7). However, I.7 states that IT budgets seem to open up again soon.

4.4.4 Comparison Between Industries

The sole presentation of the individual factors in Table 4 is not intended to suggest that each driver and barrier should be considered separately. It is clear from the results how related and correlated they can be. However, breaking them down into individual parts not only eases the understanding of AI adoption but also makes cross-industry differences and similarities apparent. This leads back to RQ3 from the first chapter.

Financial Services			Manufacturing		
Drivers					
# (FS-)	**Driver**	**TOE**	**# (M-)**	**Driver**	**TOE**
D1	Regulated data management	T	D1	Curiosity	O
D2	Scalability, reusability, flexibility	T	D2	Commitment to digitalization	O
D3	Realization of leverage effects	T	D3	Productivity and process efficiency	O
D4	Cost reduction	O	D4	Labor-intensity reduction	O
D5	Productivity and process efficiency	O	D5	Cost reduction	O
D6	Profit maximization	O	D6	Improving quality control	O
D7	Data-driven decision making	O	D7	AI ‘quick wins’	O
D8	Improving CX	E	D8	ROI through AI investment	O
D9	Competitive pressure	E	D9	Sustainability / circular economy	E
D10	Cross-company knowledge exchange	E			
D11	Achieving regulatory compliance	E			
Barriers					
# (FS-)	**Barrier**	**TOE**	**# (M-)**	**Barrier**	**TOE**
B1	Legacy IT	T	B1	AI development intertwined with physical infrastructure	T
B2	AI development tools	T	B2	Inflexibility, expensiveness of buyable AI solutions	T

Financial Services			Manufacturing		
B3	Data	T	B3	Data	T
B4	Unpredictability of AI projects	T	B4	AI model accuracy	T
B5	AI model accuracy	T	B5	'AI overengineering' (AI not the best solution)	T
B6	Explainability	T	B6	Speed of technical development	T
B7	Industry culture and risk aversion	O	B7	Cloud-isolated deployment of AI solutions	T, O
B8	Cross-functionality and agility	O	B8	Internal data privacy policies	O
B9	Leadership commitment	O	B9	Lack of standardization	O
B10	AI strategy	O	B10	Leadership commitment	O
B11	Company and workforce size	O	B11	AI deployment as a risk of operational disruption	O
B12	Know-how	O	B12	Industry failure culture	O
B13	Regulation and compliance	E	B13	Fear and change management	O
			B14	AI understanding	O
			B15	Cross-functionality and agility	O
			B16	External data privacy regulations	E
			B17	World affairs (war, Covid-19)	E
TOE elements: T - Technology, O - Organization, E - Environment					

Table 4-4 – Industry-Specific Drivers and Barriers of AI Adoption

Drivers of AI adoption. In both industries, it turns out that technological factors do not primarily drive AI adoption. There are no drivers in the manufacturing industry (the experts did not provide an explanation), while there are a few in financial services. Therefore, we can conclude that AI does not serve an end in itself. Considering both the quantity but also their scope, the lion's share of all drivers can be found in the organizational context—for both sectors. Here, it is noteworthy that AI in the manufacturing industry is driven by curiosity and commitment to digitalization. Even though these do not seem to be the strongest drivers, such aspirations in top-level management could not be found in the financial services sector. This circumstance could be related to financial services companies' corporate cultures, meaning that technology curiosity and risk aversion are presumably negatively correlated. In addition, this is an example that not only drivers among themselves but even drivers and barriers are interrelated with each other. Moreover, strong similarities in the areas of productivity, cost reduction, and efficiency can be found between the industries. These factors, in particular, can be almost completely derived from economic aspirations, as in most cases, companies will just seek to achieve a positive ROI with single AI investments (I.7). This is relatively unsurprising, as a profit motive is the nature of a private company operating in the market. Other differences include the strive for data-driven decision-making (financial services) and AI-based quality control (manufacturing). The environmental context also reveals no further similarities regarding the drivers. These differ in terms of improving CX, competitive pressure, achieving regulatory compliance, and cross-company knowledge exchange on the one hand (financial services) and sustainability on the other (manufacturing).

Barriers to AI adoption. In contrast to the drivers, significantly greater heterogeneity can be observed in the barriers for both industries. One explanation could be that challenges are not decision-based but usually stem from many sources. Here, in particular, the TOE framework provides a plausible basis for classification. Second to witness is the high volume of technological barriers in contrast to the drivers. Similarities in this context only concern the challenge of lacking data, unsatisfactory AI model accuracy, and a lack of complementary tools, albeit with divergent manifestations of this barrier. In financial services, people tendentially shy away from new types of open-source technologies for development. In manufacturing, organizations often avoid using cloud technologies to simplify solution deployment. Strong correspondences in the organizational context only exist in the fields of know-how and AI understanding, and cross-functionality and agility. However, any similarities in leadership commitment and industry culture should be drawn carefully. In the manufacturing industry, in particular, this result must be viewed from two sides: Some companies have a low level of AI adoption due to a lack of commitment. Other companies, however, are characterized—as described in the drivers—by a strong desire for curiosity and digitization at the management level. Finally, the largest industry-specific differences can be found in the environmental context. Even if high data protection regulations exist, especially in Europe, the strict requirements for financial institutions cannot be compared with any other industry. It is precisely these compliance regulations that probably represent one of the biggest hurdles for financial services providers.

4.5 Discussion

First, the findings of drivers and barriers for financial services and manufacturing are discussed and connected to the literature. Then theoretical and practical contributions, limitations, and future research follow.

4.5.1 Theorizing and Generalization of Drivers and Barriers

In the results, one could be taken aback by the almost complete absence of technical drivers. In order to find a presumed explanation, we first take a look at the theoretical cross-industry generalization of drivers in the other two contexts. Productivity, accuracy, speed, and cost reduction are industry-independent drivers and, therefore, can be interpreted as 'hard' economic factors. This finding supports the findings of general factors from Cubric (2020) that value-related incentives primarily drive AI adoption. From an outside-in perspective (i.e., mapping the generic drivers/barriers from the theoretical background onto this study's findings), other drivers can be interpreted as 'soft' determinants since these could not be found at all (e.g., increasing well-being) or not to the same extent for both industries (e.g., decision-making, sustainability, CX). This stands under the proviso that some factors (e.g., sustainability) could not yet be verified as AI adoption drivers. From an inside-out perspective (i.e., mapping this study's findings onto the generic drivers/barriers from the theoretical background), the data of this study also show that, beyond the generic drivers, there are novel industry-specific factors that cannot be generalized to other industries (e.g., regulated data management as a driver, enhanced quality control, sustainability). The strong economic orientation of the drivers discussed could represent a possible causality for the absence of technical drivers. Conceivable drivers are presumably still too detached from organizational and environmental factors at present, so that at least the conscious existence of such technical drivers is not given.

In the context of barriers, generalization becomes much more difficult. Here again, a few 'hard' barriers can be found that apply to both industries. These primarily concern the technological context (e.g., data and tool availability, AI models' accuracy) and the organizational context (e.g., required expertise and know-how, cross-functionality, agility). Nevertheless, significantly more 'soft' barriers can be found—i.e., those challenges that are not universally applicable. A look into the environmental context shows that one who wants to infer from the generic challenges (see Table 1) to those of the specific industry will likely miss out on the industry's key AI adoption determinants. The role of regulation and compliance for financial services companies, for example, cannot be generalized over all industries.

4.5.2 Contributions, Limitations and Future Research

If a company adopts AI, it has great potential for saving time and costs, thus increasing profits. Therefore, many general drivers and barriers to AI adoption have been identified in the last years (e.g., Zöll et al. (2022); Kar et al. (2021); Cubric et al. (2020) and Pumplun et al. (2019)). However, these general factors cannot indicate why AI is utilized in some industries more than others. While AI offers great possibilities in the financial service and the manufacturing industries, and some use cases are already implemented in both, they still differ significantly. Thus, we first gave an overview of existing factors in AI adoption and clustered them according to the TOE framework. We then conducted interviews to identify industry-specific drivers and barriers to AI adoption. Those were again clustered into the three TOE categories and compared across industries. It was possible to derive hard (generalizable) and soft (non-generalizable) AI adoption factors.

Our case study provides several **theoretical contributions**. First, we present an overview of the drivers and barriers to AI adoption in financial services and the manufacturing industry. Comparing these determinants resulted in a classification into soft factors (industry-specific) and hard factors (generalizable). Soft drivers in the manufacturing industry are, for example, the curiosity of companies, while an exemplary barrier in the financial service industry is their legacy IT. More generalizable factors are especially technological barriers such as available data and AI model accuracy, i.e., factors that are specific to AI. By conducting interviews, it was possible to draw a picture beyond correlated AI adoption variables and identify dynamic cause-and-effect relationships. As a result, a holistic view close to organizational structures could be given. Second, by demonstrating unique factors (soft factors) and general drivers and barriers (hard factors), we extend the TOE framework by including industry-specific determinants in technology adoption. In addition, initial evidence suggests that these drivers and barriers influence the adoption of AI to a varying degree. Therefore, we contribute to the industry-specific research of AI adoption and provide a basis for the future development of AI applications. This is likely transferable to the adoption of other technologies as well. Third, to the best of our knowledge, we are the first to combine Eisenhardt (1989) and Yin (2014) into one methodology that can guide future research. The methodology allowed a deep insight into the industries to derive the different factors of AI adoption.

Besides that, we have **practical implications** for industry stakeholders, business decision-makers, and AI executives. First, having the research designed as a multiple-case study, all results can serve as a form of anonymized, cross-company knowledge-sharing, enabling market and technology transparency in several ways. Identifying use cases and the state of AI adoption provides the necessary information to understand the current industry situation and consider the underlying context. Second, the study can serve both as a starting point for market analyses and technological investigation of AI adoption. Any industry stakeholder in financial services or manufacturing who seeks to advance their organization in the direction of AI adoption will find a detailed summary of drivers and barriers in this study. Lastly, AI specialists or consultants can benefit from understanding their customers' needs, challenges, and motivating influences behind their AI adoption journeys.

This study is also subject to **limitations**. Due to the limited number of interviews, only two industries in Germany and UK were investigated, with only three/four interviews per industry. In addition, regarding the use of the TOE framework, it is not possible to separate the drivers and barriers strictly according to the three contexts, as many factors are intertwined, e.g., having enough data is considered a technological barrier, but a data-driven culture would support getting them as part of organizational context. Pumplun et al. (2019) classified the AI readiness factor 'data' into the organizational context with the argument that data must be available and accessible within a corporate organization in order to adopt AI.

Among other thoughts, these limitations provided the basis for the following proposals for **further research**. Thus, the same research design can also be applied to other industries. Thereby, the theoretical industry-specific contributions of this study may be extended, and valuable practical insights could be given. The investigation of AI adoption in the public sector is highly recommended, as respective drivers for AI adoption may differ from economic aspirations in business. In addition, qualitative methods can focus on overcoming adoption barriers, contributing to developing corporate AI strategy frameworks. Finally, just like in previous research patterns, future studies are recommended to use quantitative methods to verify this study's findings and determine each factor's importance.

4.6 Conclusion

In this paper, we draw a holistic picture of AI adoption in the financial services and manufacturing industries. The research design enabled the elaboration of advanced and dynamic influences on AI adoption. The viewpoint has shifted from general AI readiness factors to an industry-specific perspective. The main point to emphasize here is that the driving factors for AI adoption are mostly economic aspirations. In contrast, barriers are heterogeneous, and different conclusions can be drawn for every TOE context. In the technological context, they exhibit many generic patterns across the sectors. However, the organizational context consists of a significant amount of industry-specific variations. Environmental barriers could be identified as specific and hardly generalizable. In the next step, additional industries will be added to our research to provide even more insights.

4.7 Acknowledgments

This research and development project is/was funded by the German Federal Ministry of Education and Research (BMBF) within the "Innovations for Tomorrow's Production, Services, and Work" Program (funding number 02L19C150) and implemented by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication.

5 Paper C: How Much Are Machine Assistants Worth? Willingness to Pay for Machine Learning-Based Software Testing

Title

How Much Are Machine Assistants Worth? Willingness to Pay For Machine Learning-Based Software Testing

Authors

Mehler, Maren F. and Vetter, Oliver A.

Publication Outlet

Proceedings of the Thirty-first European Conference on Information Systems (ECIS 2023), Kristiansand, Norway

Abstract

Machine Learning (ML) technologies have become the foundation of a plethora of products and services. While the economic potential of such ML-infused solutions has become irrefutable, there is still uncertainty on pricing. Currently, software testing is one area to benefit from ML services assisting in the creation of test cases; a task both complex and demanding human-like outputs. Yet, little is known on the willingness to pay of users, inhibiting the suppliers' incentive to develop suitable tools. To provide insights into desired features and willingness to pay for such ML-based tools, we perform a choice-based conjoint analysis with 119 participants in Germany. Our results show that a high level of accuracy is particularly important for users, followed by ease of use and integration into existing environments. Thus, we not only guide future developers on which attributes to prioritize but also which characteristics of ML-based services are relevant for future research.

Keywords

Software Testing, Machine Learning, Choice-based Conjoint Analysis, Willingness to Pay

5.1 Introduction

An enormous amount of novel machine learning (ML) services has recently gained traction in the market. However, pricing these services is notoriously difficult and how much users are willing to pay is uncertain, especially since it is not clear which attributes are important to the users of ML services. Various use cases of ML services exist that support users in their day-to-day tasks and support independent decision making (Berente et al., 2021; Peters et al., 2020; von Wedel & Hagist, 2022). A so far rarely explored use case of a ML-infused service supporting decision making, which shows high potential, is the application of ML in the context of software testing. Software testing is an essential part of the software development process as it ensures the quality of the software (Zhao et al., 2021). Due to the high share of human labor in form of experience and knowledge, software testing consumes a lot of time and hence leads to high costs (Enoiu et al., 2020; Guveyi et al., 2020). According to a survey among CIOs, 23% of the annual IT expenditure is spent yearly on software quality assurance. Although the costs decreased slightly in recent years (cf. 35% in 2015), approximately a quarter of the budget remains a significant percentage (Walgude & Natarajan, 2019). One approach for reducing time and costs while ensuring high quality is the use of technologies based on machine learning (in the following referred to ML-based software testing tools) (Hourani et al., 2019).

As software becomes more complex in general and the range of available software increases, the number of tests required also increases and needs to be more comprehensive. Costs and effort increase especially towards the end of the development process and even after the software has been deployed (Madera & Tomoń, 2017). In addition, safety-critical software and ML software make testing more complicated (Ebert et al., 2022). Furthermore, software companies aim to be cost-efficient and rapid in the development of software (Storey et al., 2022). These objectives reveal great potential for providers of respective tools (Durelli et al., 2019). Although some ML-based software testing tools already exist in practice and many algorithms have been presented and tested in theory (Shafiq et al., 2021), they still do not seem to be established on the market. For example, algorithms that generate test data, select test cases, or even estimate the duration of tests can be found in theory. In addition, these algorithms have been proven to be profitable based on individual use cases (Rathore & Kumar, 2021), especially since traditional testing—i.e., without the use of ML-based tools—is prone to human error (Durelli et al., 2019). Nevertheless, this emerging technology does not yet seem to have been adopted by companies, possibly because of their heterogeneous needs and individual requirements that such a tool must fulfill.

However, companies seem to have a considerable amount of budget available, which is already being spent on testing (Ebert et al., 2022). Therefore, a solution for low-cost, fast, and still high-quality testing would be ML-based tools that relieve the software testers and thus their respective companies (Shen et al., 2018). The objective of this paper is to provide first insights into the willingness to pay for ML services used in their day-to-day work. This leads to the following research questions:

1. Which attributes of ML-based software testing tools determine the willingness to pay of users?
2. How much are users willing to pay for ML-based software testing tools?

Our study can thus contribute by identifying attributes for ML-based software testing tools besides those already existing in the literature and rank them according to their importance for potential users. In addition, specific prices are linked to the attributes and the resulting product combinations. Those attributes and prices can then be generalized and used for ML services.

The remainder of the paper is structured as follows: First, an overview of willingness to pay for ML is given and the fundamentals and the most researched types of ML-based software testing in the literature are explained in section two. Then, in section three, the methodology of the conjoint analysis is presented, whereby a focus is on the literature search and Delphi study necessary for the determination of the characteristics. This is followed by the analysis of the results in section four, focusing on the results of the conjoint analysis as well as providing insights on possible user segments, which are then discussed in section five. Finally, the paper ends with a conclusion in section six.

5.2 Theoretical Background

In the following, first, an overview of previous research on the willingness to pay in the field of ML is shown. Then, ML-based software testing is introduced by presenting existing literature and defining the use case of test case generation.

5.2.1 Willingness to Pay for ML

Willingness to pay (WTP) describes the maximum amount a person is willing to pay for a product or service (Kalish & Nelson, 1991). The consumer is indifferent to buying the product for this amount (Moorthy et al., 1997). Knowing the WTP, both the sales volume and the profit margin can be maximized, and the concept is therefore used for pricing (Le Gall-Ely, 2009). Since the needs of individuals are subjective and influence WTP, determining the WTP remains challenging. Several approaches, such as experiments, direct surveys, or indirect surveys like conjoint analyses, exist to quantify consumers' willingness to pay (Breidert et al., 2006).

Before several approaches are pointed out how WTP for ML has already been addressed, it is important to define and frame the term ML first. In this paper, artificial intelligence (AI) is understood as an overarching term for ML, and ML makes decisions based on experiences (Russell & Norvig, 2021). In this paper, we restrict ourselves to the term ML.

In the field of ML, only a few papers examined the willingness to pay. One cluster addresses ML-designed products: Although people were first willing to pay more for t-shirts with an ML-generated print, the willingness to pay decreases slightly when people are informed that the print was created by a ML algorithm instead of a human (Sohn et al., 2021). H. Zhang et al. (2022) obtain a similar result: Consumers are willing to pay more for t-shirts designed by a ML algorithm. Most similar to our focus are the papers regarding the willingness to pay for ML in the medical context. For example, von Wedel and Hagist (2022) examine the experience and willingness to pay for ML-based assistance systems in the medical field. Due to the absence of adequate tools, they first identify fundamental characteristics such as the provider as key attributes for the willingness to pay. Afterward, they were able to show that the willingness to pay for such tools exists. However, due to the specific context and possible regulations in the medical context, these papers provide initial insights, and we are expanding to consider a novel use case. Other papers on WTP for ML in a broader sense are Peters et al. (2020) who can demonstrate in their study that consumers are particularly willing to pay more for transparency features in intelligent systems. In addition, Morita and Managi (2020) can show that consumers in Japan are willing to pay more for an autonomous vehicle, but the willingness remains lower than in the U.S. Finally, a research area on ML used for the calculation of the WTP exists (e.g., (Nguyen et al., 2022; Ramsey & Bergtold, 2021)).

One important method for determining WTP is a conjoint analysis. Conjoint analysis has been successfully applied for many years and was continuously improved (Green & Srinivasan, 1978). The choice-based conjoint analysis as an extension has been applied to many contexts in the lasts years like smartphone mobile operating systems (Böhm et al., 2015), platform as a service solutions (Giessmann & Stanoevska-Slabeva, 2013), content preferences of newspaper readers (Kanuri et al., 2014), web identity management systems (Roßnagel et al., 2014), augmented reality in production environments (Schuir & Teuteberg, 2021), smart meters (Albani et al., 2017), digital assistants (Ebbers et al., 2021), in-vehicle assistants (Mihale-Wilson et al., 2019). Similar to our use case, most of these papers first identify attributes, and then determine the WTP using choice-based conjoint analysis. While WTP is usually measured for consumer products, the considered product of our use case is intended for employees. Goebel et al. (2018), for example, examined the WTP of purchasing managers successfully, but noted that WTP methods are mostly applied to consumers. Since our aim is to determine the preferred features of the product, we perform a choice-based conjoint analysis with employees.

5.2.2 Machine Learning-based Software Testing

Software testing is a crucial part of the process of software development. It is a way of measuring and determining the quality of an IT system (referred to as a system under test). Depending on the project phase, single functional parts are tested as part of a unit test, or the integration of several parts, or even the entire system is tested. The primary goal is to uncover errors in order to improve and correct them. An error or bug occurs when the software does not meet the requirements or expectations of the users (Pandey et al., 2020). It is important to emphasize that a system can never be fully tested, especially only relying on human resources (Ahmed & Zamli, 2011). Therefore, it is recommended to perform diverse and as many tests as possible (Segura & Zhou, 2018). To support the software testers in their day-to-day work, numerous ML-based software testing tools have been developed in theory (e.g., (Dejaeger et al., 2013; Kesri et al., 2021; Tahvili et al., 2018)). In the context of this paper, ML-based software testing can support manual testing, but especially improve automated testing.

Both in literature and practice, two concepts should be distinguished (Gezici & Tarhan, 2022): Software testing, where ML provides support—this is the focus of this paper—and the testing of ML algorithms. The latter may also be facilitated by ML and could utilize the same algorithms, but the use case is highly specific and could lead to confusion in the survey conducted as part of this paper. In the following, ML-based software testing is thus referred to as the testing of ordinary software with the support of ML.

There is neither a single nor a standardized solution for ML-based software testing since the variety of different tests and testing approaches is matched by the variety of algorithms and tools that exist or have been proposed in the literature. They differ, for example, in terms of the applied AI algorithms—neural networks, clustering such as kNN or decision trees are commonly utilized (Lima et al., 2020). In the context of this paper, the available ML-based software testing tools are categorized according to their areas of application (Durelli et al., 2019; Hourani et al., 2019; Shafiq et al., 2021). The following exist:

- Test Case Generation—Generation of test cases from requirements or code using Natural Language Processing (NLP) (Dadkhah et al., 2020; Kesri et al., 2021; Memon et al., 2001).
- Test Data Generation—Generation of test data, meaning the automatic identification of input data for test cases, e.g., through specific ML algorithms like backtracking, AI planning, or swarm intelligence (Gupta et al., 2004; Rabbi et al., 2018; Xing et al., 2014).

- Test Oracle Generation—Test oracles are the set of program results depending on the input data (Barr et al., 2015); these can also be extracted automatically using classification, support vector machines, or neural networks (Gholami et al., 2018; Khatibsyarbini et al., 2021).
- Test Case Selection/ Prioritization—To save time and costs, it is convenient to perform only certain tests or to begin with the most important ones. The selection or prioritization can be supported by ML algorithms in the areas of classification, clustering, or swarm intelligence (Kazmi et al., 2017; Khatibsyarbini et al., 2021; Pan et al., 2022).
- Test Estimation—ML can also be integrated into secondary tasks such as estimating the remaining duration of tests, e.g., using regression. This can support managers in their planning of the remaining software project or test case selection/prioritization (Tahvili et al., 2018).
- Software Fault Prediction/ Bug Prediction/ Defect Prediction—All terms are used synonymously and describe the prediction in which modules of the software errors are likely. Algorithms such as dimensionality reduction, Naive Bayes, or neural networks are applied here (Catal & Diri, 2009; Dejaeger et al., 2013; Gondra, 2008; Z. Li et al., 2018).

Throughout this paper, the focus is primarily on test case generation, since this use case is a common task for software developers and software testers. Additionally, the adoption of ML for test case generation is considered to be the most suitable use case by the experts in our—later described in detail—Delphi study. Therefore, this use case will be presented in more detail.

It is extremely time-consuming to manually generate test cases from requirements, as these often do not match a given standard since they are written by several domain experts. Thus, natural language processing (NLP) algorithms can support software testers by reducing time and costs, as well as producing results without human errors (Kesri et al., 2021). In addition, test cases can be derived directly from code. Since many dependencies exist between the system modules, the generation of test cases can become very complex. Therefore, a form of knowledge management is often applied during the generation of test cases (Dadkhah et al., 2020; Kesri et al., 2021). Furthermore, automated planning for graphical user interfaces was utilized for test case generation (Memon et al., 2001).

ML appears to have arrived in software engineering research, especially in software testing (Shafiq et al., 2021) and numerous opportunities exist to reduce software testing costs and time and minimize errors. However, although many tools assist programmers, human testers are not obsolete and will not be replaced. Many tasks, such as defining the test objectives or analyzing the test results, remain the responsibility of individuals, which means that the focus of ML-based software testing tools is on providing support (Itkin et al., 2019; T. M. King et al., 2019; Marselis, 2020).

Although a substantial amount of literature on potential algorithms already exists and an increasing number of tools are being developed, the economic aspects of ML-based software testing have hardly been examined. Most papers present a new algorithm that is evaluated based on ML-specific criteria such as accuracy (e.g., (Dejaeger et al., 2013; Kesri et al., 2021; Tahvili et al., 2018)). Little literature can be found on whether ML-based software testing is profitable. Rathore and Kumar (2021) perform a cost-benefit analysis on their algorithm and can demonstrate for almost all datasets that the application of their algorithm for fault prediction can reduce software testing costs. Herbold (2021) addresses the costs and profit of software defect prediction in his paper. He presents a cost model which includes different aspects such as the costs for software defects which are discovered after the release of the software. To the best of our knowledge, no research goes beyond the focus on costs and investigates the willingness to pay for such solutions.

5.3 Methodology

To answer the research questions, we conducted a conjoint analysis, as such analysis is used "to measure the joint effects of a set of independent variables on the ordering of a dependent variable" (Green & Rao, 1971, p. 355). This involves determining the preferences of (future) users and generating predictions about purchasing behavior (Green & Rao, 1971). In this paper, we apply a choice-based conjoint analysis, originally by Louviere and Woodworth (1983), which differs from the traditional conjoint analysis by presenting respondents with different product combinations several times in a row and asking the respondents to decide between them (see Figure 5-1) and thus selecting the product with the highest personal utility (Ebbers et al., 2021). This results in more realistic data through the questioning of potential users, since the method of questioning leads to real preferences (Mihale-Wilson et al., 2019).

Before conducting a choice-based conjoint analysis, the product attributes and their characteristics (levels) must be defined. To determine the attributes and levels, a two-step approach is chosen: First, a structured literature search according to vom Brocke et al. (2009) is conducted, followed by a Delphi study according to Skinner et al. (2015) to determine additional attributes and levels and eventually select them.

5.3.1 Literature Review

The literature search was conducted during the first quarter of 2022. In the **first phase** ("definition of review scope") we follow Cooper's taxonomy as suggested by vom Brocke et al. (2009) (see Table 5-1). We *focus* on the *research outcome* of the considered papers. Our *goal* is the *integration* of our results. We take a *neutral perspective*. The *coverage* is *representative* and the *organization* is *conceptual*. Our *audiences* are *specialized scholars and practitioners* (Cooper, 1988; vom Brocke et al., 2009).

<table>
<tr><td>Characteristic</td><td colspan="12">Categories</td></tr>
<tr><td>Focus</td><td colspan="3">Research outcome</td><td colspan="3">Research methods</td><td colspan="3">Theories</td><td colspan="3">Applications</td></tr>
<tr><td>Goal</td><td colspan="4">Integration</td><td colspan="4">Criticism</td><td colspan="4">Central issues</td></tr>
<tr><td>Organization</td><td colspan="4">Historical</td><td colspan="4">Conceptual</td><td colspan="4">Methodological</td></tr>
<tr><td>Perspective</td><td colspan="6">Neutral Perspective</td><td colspan="6">Espousal of position</td></tr>
<tr><td>Audience</td><td colspan="3">Specialized scholars</td><td colspan="3">General scholars</td><td colspan="3">Practitioners/ politicians</td><td colspan="3">General public</td></tr>
<tr><td>Coverage</td><td colspan="3">Exhaustive</td><td colspan="3">Exhaustive and selective</td><td colspan="3">Representative</td><td colspan="3">Central/ pivotal</td></tr>
</table>

Table 5-1 – Taxonomy of Literature Review Following Cooper (1988)

For the **second phase** ("conceptualization of topic") we got an overview of the different concepts of ML-based software testing using a general Google search and Google Scholar. We started by using the terms "artificial intelligence" and "machine learning" as synonyms in combination with "software testing". We quickly realized that there is not one, but many different approaches on the one hand regarding the ML algorithms used, on the other hand regarding the different possible applications of ML in software testing. Overviews provided by Durelli et al. (2019), Hourani, Hammad and Lafi (2019) and Shafiq (2021) were used for the background section of ML-based software testing. Our search also revealed that besides “software testing”, the terms “quality assurance” and “quality management” are often used. However, the combination of “quality assurance” and “artificial intelligence” or “machine learning” also results in many articles on quality assurance in manufacturing using AI. To limit the search, we added “software” to our search term.

Finally, the search string “((“software testing” OR “quality assurance” OR “quality management”) AND (“artificial intelligence” OR “machine learning”) AND “software”)” resulted, which is used in the **third phase** (“literature search”). The search term gives 827 results in Web of Science and 158 results in Ebscohost Business Source Premier, respectively. After removing duplicates, 953 papers are left. In the next step, the titles and abstracts of those papers were analyzed. In this process, we excluded papers with 1) very specific use cases such as medicine, farming, manufacturing, etc. as those mainly utilize ML algorithms to increase the quality in general, not software quality. In addition, we excluded papers that 2) focus on only the testing of ML algorithms as those are—as already argued in the section on ML-based software testing—out of our scope. This resulted in 468 peer-reviewed papers. While reading these papers we noticed that they primarily focus on the application, implementation, or benchmarking of a specific or new algorithm (e.g., (Bouktif et al., 2014; Rabbi et al., 2018; Tahvili et al., 2018)) and thus 3) are not focusing on economic aspects and user demands of ML-based software testing. After excluding those papers, 18 papers are left; mainly literature reviews.

In the **fourth phase** (“literature analysis and synthesis”), all remaining papers were read and analyzed for possible attributes for the conjoint analysis. However, the papers rarely outlined attributes, which might be essential for ML-based software testing tools. The following features could be extracted:

- Accuracy, as a measure of the correctness of the tool (Briand, 2008)
- Additional training data needed to implement the tool successfully in the company (Arora et al., 2015; Durelli et al., 2019)
- Ease of use of the tool, e.g., with a user-friendly interface (Arora et al., 2015; Dejaeger et al., 2013)
- Explainability, as the ability to explain the predictions (Bouktif et al., 2014; Schieferdecker, 2020)
- Possibility for users to intervene during the calculations of the tool (Durelli et al., 2019)
- ML skills needed to use the tool (Poth et al., 2019)

The **last phase** (“research agenda”) is addressed in the discussion (see section 5.5).

5.3.2 Delphi Study

However, the above-mentioned factors are not sufficient to perform a conjoint analysis as levels and prices are missing. Therefore, a Delphi study based on Skinner et al. (2015) was conducted in June and July 2022. A Delphi study is useful when the opinion of a panel of independent experts is needed on a specific topic, as it is required to determine attributes, levels, and prices. The fifteen experts selected were software testers, in particular test/ quality assurance automation engineers including six experts with decision-making responsibilities. Half of the fifteen experts have between three months and five years of experience with ML-based software testing, particularly in the areas of test case generation, test data generation, and fault/bug/defect prediction. The other half were experienced software testers without specific knowledge of ML. This gives a good mix of both experts in the field of ML-based software testing and experts in general software testing, but inexperienced users regarding ML-based software testing.

In sequential rounds, the experts were asked via emails to participate in an online survey about features, their importance, possible levels, and prices. For the specific case of test case generation—as this was most familiar to the experts—attributes such as accuracy, adaptability, cost-saving, community, documentation, ease of use, explainability, flexibility, integration, reliability, security, and test case coverage were listed. The three which ranked highest (~10% of all mentioned attributes following Skinner (2015)) were accuracy, integration, and ease of use. The derived levels are shown in Table 5-2. Willingness to pay for one license per month indicated by the participants ranged from €0.99 to €2000. Here the median of €150 was chosen and the prices of €50, €150, and €250 were derived. In addition to the statements of the experts, pricing of existing tools was referred to for comparison. The tools available on the market are quite comprehensive and vary in terms of what they offer, so only a rough price range can be given here. Since ML-based software testing tools are still new, many companies do not state their pricing on the website, possibly because they are determining their customers' willingness to pay. Thus, for example, Github offers the open source tool Github Copilot, which is available for business users for $19 per month per user. Here, the focus is on the completion of code in particular (GitHub Copilot, 2023). Perfecto offers a license starting from $125 per month which provides in particular automatic debugging (Perfecto, 2023). Sofy offers for $549 per month, among other things, test case selection/prioritization (Sofy, 2023). Thus, the pricing of existing tools confirms the prices from the experts. In the last round of the Delphi study, the consensus of the participants was confirmed.

Attributes	Attribute Levels
Accuracy	Low (90%)
	Moderate (95%)
	High (99%)
Integration	The tool is **separated** from the programming environment.
	The tool is **integrated** into the programming environment.
Ease of Use	The tool takes some time to get used to and requires some familiarization.
	The tool is simple to use with a user-friendly interface.
Price per license per month	€50
	€150
	€250

Table 5-2 – Attributes and Attribute Levels for the Conjoint Analysis

5.3.3 Conjoint Analysis

After determining the attributes and levels through the iterative questioning of experts in the Delphi study, the choice-based conjoint analysis was designed using the software tool Conjointly (Conjointly, 2022) and integrated into an online survey. The survey is structured as follows: First, an introduction to ML-based software testing and in particular to the use case of test case generation was presented with a focus on the attributes and their levels (see Table 5-2). Then, the participants were randomly assigned 12 combinations according to the scheme in Figure 5-1 with the task of selecting the best combination for them or none at all. This “none” option is a special design choice that leads to a balanced response behavior (Gensler et al., 2012). Finally, we collected demographic data such as age, gender, and software testing experience, as well as company size.

After developing the questionnaire, we conducted a pretest among IS researchers independent of this project for clarity, framing, and the time needed to participate. Subsequently, small details were added for better understanding, such as a visual representation of the levels. Finally, the participants were approached by a research institute, as the target group is very specific. It was ensured that the participants had experience in software testing.

Integration	The tool is **integrated** into the programming environment	The tool is **integrated** into the programming environment	The tool is **separated** from the programming environment	None of them
Accuracy	Low (90%)	High (99%)	Moderate (95%)	
Ease of use	The tool is **simple to use** with a user-friendly interface	The tool **takes some time to get used to** and requires some familiarization	The tool is **simple to use** with a user-friendly interface	
Price per license per month	€150	€250	€50	

Figure 5-1 – Sample Choices for Conjoint Analysis

A total of 119 software testers participated in the survey conducted in September 2022. Participations that seemed unrealistic due to a short response time and duplicates were eliminated directly by the survey tool. On average, the duration of the survey took a little over 7 minutes. By recruiting the participants through a market research institute, it was ensured that all the participants work in the quality assurance unit of an IT department, i.e., software testing, in Germany, which is important for the price range.

		#	%
Gender	Male	66	55.5
	Female	53	44.5
	Divers	0	0.0
Age	18-25	10	8.4
	26-35	50	42.0
	36-45	32	26.9
	46-55	19	16.0
	56-65	6	5.0
	Over 66	2	1.7
Experience Software Testing in years	Less than 1	18	15.1
	1-2	31	26.1
	3-5	38	32.0
	More than 6	32	26.9

		#	%
Number of Employees	Less than 100	29	24.4
	100-499	36	30.3
	500-999	20	16.8
	1000-2499	21	17.6
	2500-9999	7	5.9
	Over 10000	6	5.0
Experience Test Case Generation in years	Less than 1	29	24.4
	1-2	45	37.8
	3-5	22	18.5
	More than 6	23	19.3
Years employed in current company	Less than 1	9	7.6
	1-2	34	28.6
	3-5	36	30.3
	More than 6	40	33.6

Table 5-3 – Demographic Data and Control Variables

The gender distribution is balanced with 55.5% of participants identifying as male, and the remaining as female. Almost half (42%) of the participants are between 26 and 35 years old; a quarter (26.9%) are in the range of 36-46 years and the remaining participants are split among the other age groups (see Table 5-3). A third (32%) of the participants have three to five years of experience in the field of software testing; only 15% have less than one year of experience. As expected, experience in the area of test case generation is slightly lower with about 40% having one to two years of experience and about a quarter (24.4%) having less than one year of experience. The respondents have also been employed in their current company for quite a long time (63.9% at least three years). Regarding the size of the company, expressed as the number of employees, no clear trend can be observed; only large (> 2,500 employees) and very large companies (> 10,000 employees) are hardly represented.

5.4 Results

Figure 5-2 shows the relative importance of the attributes. Besides the most important attribute, price per license per month, participants tend to have a special interest in the accuracy of the tool. Integration and ease of use seem to be similarly relevant to the participants.

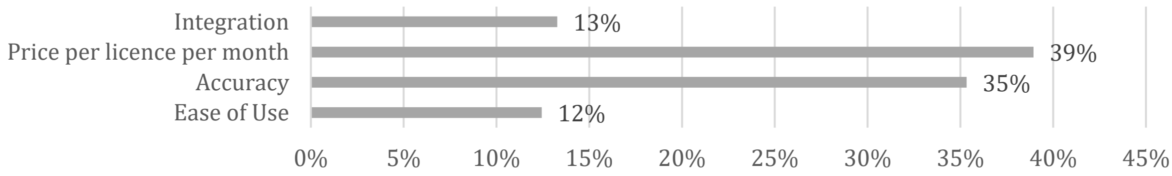

Figure 5-2 – Relative Importance of the Attributes

The respective level part worth (often referred to as level values, conjoint analysis utilities, or attribute importance scores) are shown in Figure 5-3, which reflect the preferences of an average consumer. More preferred levels receive a higher rating than less preferred levels. The individual levels per attribute add up to zero. The most preferred level is an accuracy of 99%; the second most is a price of €50 per license per month. The least preferred level is an accuracy of 90% and the second least preferred is a price of €250.

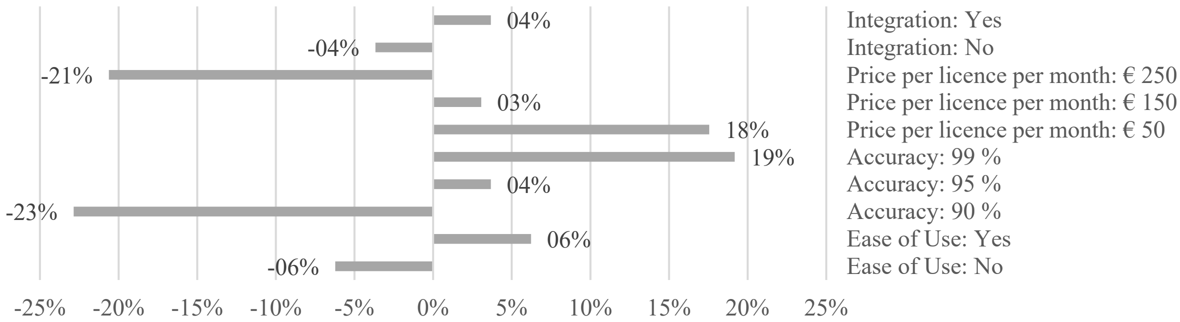

Figure 5-3 – Level Part Worth

Next, Figure 5-4 gives the distribution of preferences for the respective levels. For both integration and ease of use, respondents are nearly indifferent overall with about half selecting “yes” and half “no”. In addition, half of the participants chose €50 per license per month as the preferred level, and also half 99% as the most preferred accuracy.

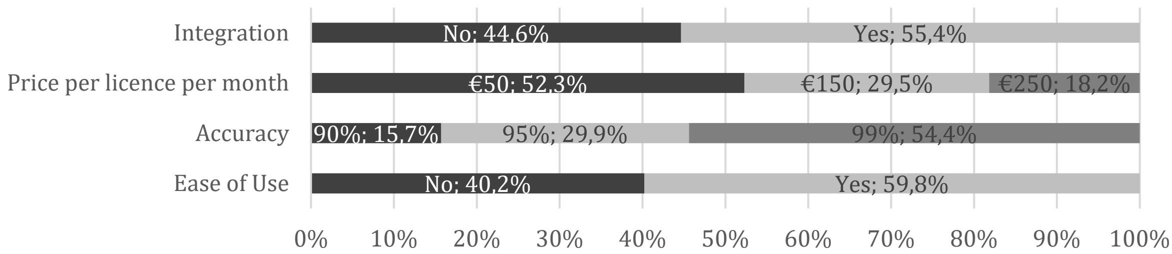

Figure 5-4 – Distribution of Preferences for the Respective Levels

Combining the attributes and respective levels into products, it can be observed that the most preferred product has the following attributes: A simple-to-use tool with a user-friendly interface, an accuracy of 99%, and possible integration into the respective programming environment for

€50 per license per month. This combination could generally be described as the "best" product, since all attributes achieve the highest level, while the price is the lowest.

Based on this result, the marginal willingness to pay is determined. It describes the amount that participants are willing to pay for a certain level of an attribute, i.e., the additional amount they are willing to pay to switch from level A to a superior level B (see Figure 5-5).

Figure 5-5 – Marginal Willingness to Pay

To the median user, an easy-to-use tool is worth as much as a price reduction of €24.39. To a median person, an increase in accuracy from 90% to 95% or 99% is worth as much as a decrease in price from €85.26 and €117.87, respectively. And to the median person, the possible integration of the tool into a programming environment is worth as much as a reduction of the price by €17.21.

Finally, the marginal willingness to pay will be examined by dividing the participants into customer segments according to the demographic data gender, age, experience in software testing and the control variable size of the company as these provide the most interesting results. However, the findings must be interpreted with caution, as they are derived from a smaller number of data points. Figure 5-6 shows the marginal willingness to pay according to gender. The baseline willingness to pay is omitted, as it is always zero. Interestingly, women are willing to pay more for ease of use, but less for integration. They are also willing to pay almost twice as much as men for 99% accuracy.

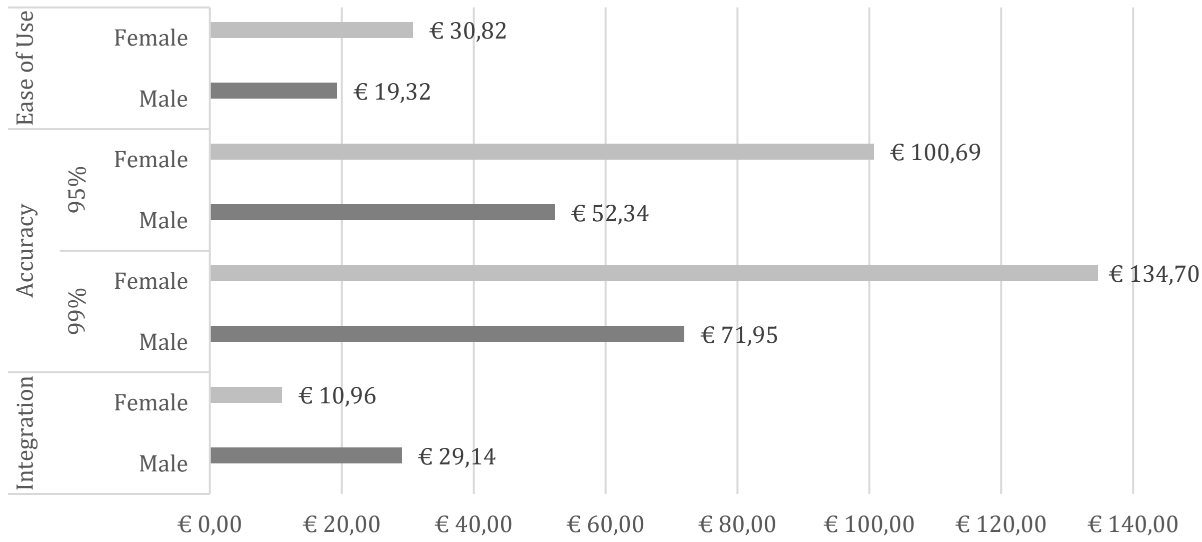

Figure 5-6 – Marginal Willingness to Pay According to Gender

Likewise, the segmentation by age (see Figure 5-7) shows large differences. The age group between 36 and 45 is particularly willing to pay a great amount for high accuracy. When it comes to integration, it can be seen that people over 56—participants between 56 and 65 and over 65 were grouped as the small number is not representative—have a negative willingness to pay.

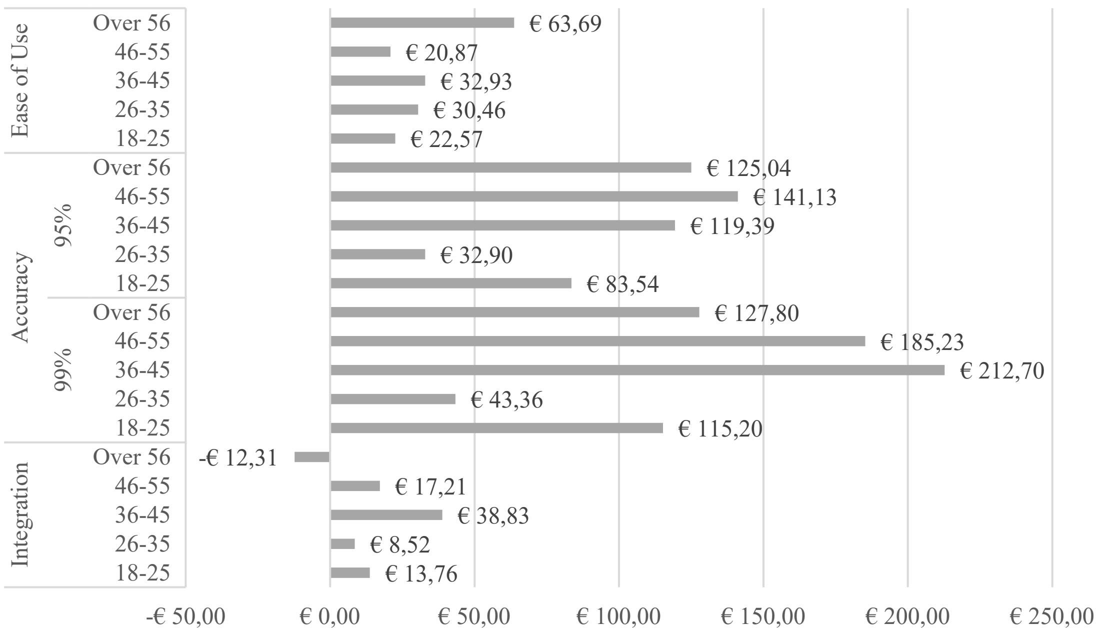

Figure 5-7 – Marginal Willingness to Pay According to Age

Lastly, the segmentation according to the years of experience and to the number of employees cannot be included as a figure due to page limitations, but show interesting results. First, regarding the experience: It shows that experienced employees in particular are willing to pay more for integration (€73.51 compared to around €10). With more experience, the WTP for accuracy increases and the WTP for ease of use is balanced throughout the years of experience. Second, regarding the size of the company: Software testers from smaller companies are willing to pay less for the attributes, except for companies with 100 or less employees, here the willingness increases. The WTP increases with the size of the company with two exceptions: Employees of companies with between 2500 and 9999 employees are willing to pay the most for 99% and 95% accuracy with €386.25 and €345.92, respectively. Small companies with 100 or fewer employees would pay least for integration with only €8.74, and the WTP for integration also drops to €17.21 for companies with 1000-2499 employees.

5.5 Discussion

Although software testing is an essential part of software development and a large budget is spent on it, many organizations seem hesitant to develop suitable ML-based software testing solutions. The adoption of such a ML-based tool would not only reduce the cost but also decrease the time required and prevent human errors. Therefore, this paper examined the ideal design of ML-based software testing tools by evaluating the willingness to pay of potential users using a choice-based conjoint analysis. A two-step process was adopted to determine the required attributes, levels, and prices: First, a structured literature review and subsequently, a Delphi study with experts in the field of ML-based software testing.

This approach provides several **theoretical contributions**. First, many concepts and algorithms utilizing ML exist in the literature to support software testers and developers in their daily work. However, hardly any economic aspects or characteristics that are essential when selling these tools to users are considered (Barney et al., 2012). Some attributes such as accuracy (Briand, 2008), ease of use (Arora et al., 2015; Dejaeger et al., 2013), or explainability (Bouktif et al., 2014; Schieferdecker, 2020) were gathered as part of the literature review. In addition, we determined new attributes, as well as their levels and prices as part of our Delphi study. In particular, accuracy was identified as an important attribute by the experts. It is not only used by most of the papers from the literature review to evaluate the quality of their algorithm or tool (e.g., (Dejaeger et al., 2013; Kesri et al., 2021; Tahvili et al., 2018)), but is also identified by the experts as a key attribute. Similarly, ease of use of the tool was rated as essential by the experts, thus this attribute from the literature is confirmed in practice. The additional attributes identified in the literature, such as ML skills or the extension of the tool with own test data, however, were rated as having less influence on the purchasing behavior of a user in the Delphi study. Instead new attributes such as flexibility, integration, and reliability were identified. Peters et al. (2020), for example, showed that the attribute transparency is important for intelligent systems, which was also identified in the Delphi study. Thus, these attributes are not limited to just software testing tools but can also be applied to other ML-based tools, especially accuracy, which is a typical measure for ML. In addition, the attribute ease of use is generalizable for tools, since they should especially be easy to use in order to be adopted. Furthermore, the results of the conjoint analysis demonstrate the users' view on ML-based software testing tools which results in price as the most important attribute for the participants. Besides price, accuracy is considered to be the most essential attribute for potential users. The participants are willing to pay up to €120 per license per month more for an increase in accuracy from 90% to 99%. Integration and ease of use are also perceived by the participants as being relevant as the second and third most important attributes and they both receive a marginal willingness to pay of around €20. This implies that future research should focus on explainability. Although the high willingness to pay indicates the users' preference for exceptionally high accuracy, the increase from 95% to 99% is very costly in the development. This leads to the question of whether other attributes should be emphasized instead of such a high increase in accuracy during the development. From this some **practical contributions** follow. First, as only a limited number of tools exist on the market, but a lot of research is already conducted on how to develop ML-based software testing tools, we encourage the implementation of these tools from theory into practice, especially as many companies have a relative high budget available. Due to the focus of the participants on price, a specific price strategy by gaining market shares via initially low prices and bounding users by lock-in effects can be recommended. The price can then be increased over time and, above all, the willingness to pay of different user types can be skimmed off by offering different product combinations. In addition, it is possible to integrate additional use cases to test case generation into one tool or platform. Second, users can be segmented especially according to number of employees as licenses for ML-based software tools are mainly sold to companies. The analysis of user segments revealed a negative willingness to pay for the attribute integration among the oldest participants. Thus, the integration should be optional, i.e., users should be able to choose whether they want to integrate the ML-based tool into their existing programming environment.

However, our contributions are also subject to **limitations**, which must be considered when identifying and adopting the results or for the design of similar studies. First, the willingness to pay determined in the conjoint analysis depends on the attributes and levels selected and presented. Although these have been developed and tested by experts, an additional attribute or level can alter the analysis. In addition, the marginal willingness to pay is only within the range of the predetermined prices. However, different prices would only change the level of the willingness to pay and should not lead to a different result. Therefore, the conclusions drawn only apply to the selected combination. Second, since there are currently not many ML-based software testing tools on the market, it can also be assumed that most participants in the conjoint analysis have little or no experience so far. Hence, it can only be derived that the attributes in the order accuracy, integration, and ease of use are important to our participants. This might also change over time as users become familiar with ML-based tools. Finally, the focus on price must be interpreted with caution, since the respondents usually get the tools paid by their employers. Thus, if the prices of the tools are eventually within the budget, the exact price is presumably not important to a single user, but to the company.

Therefore, we can derive **further research** from the contributions and limitations. We want to focus on three: First, examining the WTP of other ML-based services, and thus other attributes, could lead to additional insights. Especially comparing the WTP of traditional services and ML-based services could lead to fascinating results. Second, it would be interesting to offer users a real tool to determine their willingness to pay. Third, to the best of our knowledge, we are one of the first to conduct choice-based conjoint analysis in a corporate context. This could lead to develop new methods for future research.

5.6 Conclusion

In summary, it was shown that machine learning has entered the software testing literature and offers great potential in practice. Not only can time and costs be saved through the automatic development of test cases, but human errors can also be reduced. In particular, high accuracy, an easy-to-use tool, and integration into the programming environment are important to users and lead to an increase in willingness to pay. The most preferred product combination for the users is 99% accuracy, high usability for example through a user-friendly interface, and the ability to integrate the tool into an existing programming environment for a price of €50 per month for a license. Other interesting attributes such as transparency or availability of documentation were identified which might be relevant for other ML-based services. However, as such tools and especially ML-based software testing tools are hardly available in practice, it will be worth developing them with a focus on accuracy, ease of use, and integration.

5.7 Acknowledgment

This research and development project is/was funded by the German Federal Ministry of Education and Research (BMBF) within the "Innovations for Tomorrow's Production, Services, and Work" Program (funding number 02L19C150) and implemented by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication.

6 Paper D: Performance vs. Purpose: Generative AI Enhances Task Performance but Reduces Meaningfulness in Programming

Title

Performance vs. Purpose: Generative AI Enhances Task Performance but Reduces Meaningfulness in Programming

Authors

Mehler, Maren F. and Krautter, Kai

Publication Outlet

Proceedings of the Thirty-Second European Conference on Information Systems (ECIS 2024), Paphos, Cyprus

Abstract

Generative Artificial Intelligence (GenAI) has become widespread in daily work but present novel challenges for users as previously meaningful tasks can now be completed by GenAI. This study examines the impact of ChatGPT on task performance and perceived meaningfulness in two programming tasks. In an online experiment (n=161) assigning participants to coding or debugging tasks, with and without ChatGPT assistance, we found that using ChatGPT improved task performance, partially because the supported tasks are less difficult. However, using ChatGPT resulted in lower perceived meaningfulness, partly because participants considered the tasks less effortful. Notably, both tasks exhibited slightly different results, indicating that contextual factors may amplify or mitigate the effects. This study emphasizes the dual nature of GenAI integration, balancing enhanced performance with psychological impacts on users. Our findings offer insights for organizations and developers on integrating GenAI, highlighting the importance of incorporating efficiency gains with the meaningfulness of human work.

Keywords

Meaningfulness, Task Performance, Generative AI, ChatGPT, Online Experiment

6.1 Introduction

In an era where technology has become an integral part of our daily lives, artificial intelligence (AI) facilitates various aspects of work (Berente et al., 2021; Brynjolfsson & Mitchell, 2017; Russell & Norvig, 2021). For example, AI is expected to create new roles with new tasks requiring skills such as creativity, problem-solving, or empathy (Bryant, 2023). As AI evolves, advances in Generative AI (GenAI), especially in Large Language Models (LLMs) over the course of the last year, have resulted in tools such as ChatGPT, which are expanding the potential applications and fostering a new era of collaboration between AI and humans. In areas like customer service, where ChatGPT can provide responses, in marketing, where GenAI can be used to produce content such as images with Dall-E 3 (Northwest Executive Education, 2023; OpenAI, 2023a) or GitHub Copilot in supporting coders in their daily work (GitHub Blog, 2023) new possibilities arise. So far, AI has been effective at performing routine tasks (e.g., translating (DeepL, 2023), freeing up human capacity for more engaging, creative and meaningful tasks. However, the introduction of content creation tools—GenAI—using LLMs such as ChatGPT and image generation tools like Dall-E 3 is changing the landscape and raising intriguing questions about the extent to which AI can now handle tasks that were once considered the exclusive domain of human creativity (Dwivedi et al., 2023).

Thus, the main objective of this paper is to examine whether individuals (i) improve their performance in complex activities with GenAI assistance (hereafter referred to as *task performance*) and (ii) how the usage of (generative) AI impacts how they feel about their work (hereafter referred to as *perceived meaningfulness*). Specifically, we are interested in how employees perceive their work as meaningful, given that finding passion and meaningfulness in their work is a top priority for many employees (Cech, 2021; Jachimowicz & Weisman, 2022). Academic research has supported the notion that finding one's work meaningful can increase motivation and, consequently, job performance (Allan, 2017; Jachimowicz et al., 2018). To better understand the interaction between AI and human employees, it is thus crucial not only to consider how people perform at work but also how they feel about their work.

We test our research questions in the context of programming—an area within corporate organizations that stands to derive significant advantages from the assistance of GenAI. In addition, a fusion of technical expertise and creative problem-solving is paramount in this domain. Specifically, ChatGPT can aid in various aspects, such as code generation, error detection, and testing (Perkel, 2023). While many possibilities to adopt ChatGPT and ease work exist (Dwivedi et al., 2023), ChatGPT cannot serve as a substitute for the primary contributions of programmers. Still, it can effectively streamline and enhance their work processes (Larson, 2023). Therefore, we examine whether AI assistance can increase task performance in programming tasks and whether it affects how people feel about their work (i.e., how meaningful they perceive their work). In an online experiment we asked 161 participants to either complete coding or debugging tasks with or without ChatGPT's assistance. We observed that partly due to less difficulty higher task performance could be attained with the assistance of ChatGPT. Conversely, the use of ChatGPT resulted in partially viewing the tasks as less effortful and thus lower perceived meaningfulness was found among participants. However, as both tasks results are slightly different, various contextual factors may influence our observed effects. Consequently, our results suggest that ChatGPT's assistance influences individuals' perceptions of their work, a factor that should be taken into consideration in future decisions regarding human-AI collaboration.

The paper is structured as follows: In the next chapter, we outline the background and address the development of AI into GenAI. We also focus on how task performance and perceived meaningfulness have been measured and used in prior research. In the third chapter, we derive our hypotheses for our online experiments. The fourth chapter presents our methodology, where we describe the design of the experiment and the data collection process. This is followed by the analysis of the results in the fifth chapter, and a discussion concludes this paper with the findings, our contributions, limitations, and future research in the last chapter.

6.2 Theoretical Background and Hypotheses Development

This chapter is divided into two parts. First, we provide a definition and overview of AI, focusing on the characteristics of GenAI and ChatGPT, as well as the influence of GenAI on programming. Then, we will present an overview of the measurements of work engagement: task performance, i.e., how people perform at work, and perceived meaningfulness, i.e., how they feel about their work. Additionally, we point out previous work on both aspects in combination with ChatGPT.

6.2.1 Artificial Intelligence

In the field of AI, the emergence of GenAI, exemplified by systems like ChatGPT, has introduced in a paradigm shift in the perception and interaction with AI technologies. Previously, AI and its subcategory, machine learning (ML), were learning algorithms making decisions or predictions based on recognized patterns in the data (Brynjolfsson & Mitchell, 2017; Mitchell, 1997; Russell & Norvig, 2021). However, with GenAI and LLMs, tools such as ChatGPT, Bing Chat, or, specifically for programming tasks, GitHub Copilot, it became possible to generate new data from identified patterns rather than solely analyzing the data (Dwivedi et al., 2023; Teubner et al., 2023). Tasks that were once assumed to be exclusively within the realm of human capability were already being replaced before the advent of GenAI (Bankins & Formosa, 2020), and this trend may further intensify now.

Prior research on human-AI collaboration has primarily operated under the assumption that AI can make decisions but not generate content (e.g., Fügener et al., 2021; Jussupow et al., 2021). It has been demonstrated that when the AI is correct, it can enhance human performance (Boyacı et al., 2023; Fügener et al., 2021), emphasizing the potential benefits of humans and AI working together on tasks. However, this research also indicates that humans tend to prefer challenging tasks and are hesitant to delegate them to AI systems, even though this delegation could improve performance (Fügener et al., 2022). These initial findings suggest that how humans feel about their work is often more important to them than how they perform.

With regard to programming, ChatGPT or more specific tools such as GitHub Copilot offer programmers the means to simplify their daily work with high-quality output (GitHub Copilot, 2023; Perkel, 2023). Programmers can now use AI to draft code, debug it, run tests, or receive explanations (E. Chen et al., 2023; Liu et al., 2017; Surameery & Shakor, 2023). As a result, these new possibilities are reshaping the way programmers work, which, in turn, impacts their performance and how they perceive the meaningfulness of their work. Consequently, ChatGPT has the potential to fundamentally transform how humans interact with technology (Haleem et al., 2022).

6.2.2 Engaging in Work

Engagement in daily work is twofold. On the one hand, it needs to be measured by employers which is mainly done through performance measurements. On the other hand, the work is assessed by how employees feel towards it—finding meaning in their work.

6.2.2.1 Performance

Task performance can be defined as the effectiveness with which a person carries out an activity, often pertaining to a task related to their job (Borman & Motowidlo, 1993, 1997). These activities have a direct connection to organizational outcomes, as they either transform resources into products and services or support such processes (Motowidlo & Scotter, 1994). Performance can generally encompass various aspects, such as the quality, quantity, and relevance of service (Makki & Abid, 2017). The specific interpretation of performance may vary based on the particular field or area of interest. In our context, we define task performance based on the quality of execution of a programming assignment. This can be subjectively evaluated by the person performing the assignment (Goodman & Svyantek, 1999), but it can also be assessed more objectively by analyzing the individual task output (see Benlian, 2015), which in our case refers to the final code snippets. Since measuring performance is very complex (Ramos-Villagrasa et al., 2019), we decided to use both.

In AI research, performance measurements, such as accuracy, are frequently used to demonstrate the effectiveness of algorithms (Russell & Norvig, 2021), similar to research in human-AI collaboration, where performance is measured to indicate its improvement when AI is employed (e.g., Boyacı et al., 2023; Fügener et al., 2021). Noy and Zhang (2023) have presented a working paper demonstrating that productivity can be enhanced when ChatGPT is used. They also find evidence that ChatGPT can boost job satisfaction. Gilardi et al. (2023) went one step further and demonstrated that ChatGPT outperforms crowd workers by an average of 25 percentage points for annotation tasks. Nonetheless, the measurement of performance using GenAI is still in its early stages. For instance, Chen et al. (2023) conducted an empirical study for evaluating text quality, and they identified asking ChatGPT for an explicit score as the most reliable and effective one.

Assessing how well a person performs in a programming assignment can be done by evaluating code quality. Various metrics, such as complexity, can be used for this purpose (Baggen et al., 2012). More specifically, there are many metrics falling into several categories. For example, the code's degree of complexity can be measured through code comprehensibility, while the absence of errors can be assessed through software bug prediction (Nuñez-Varela et al., 2017). Additionally, coding standards and code reviews can enhance quality (Boogerd & Moonen, 2008; Kononenko et al., 2016; X. Li & Prasad, 2005; Stegeman et al., 2014). GenAI opens up new possibilities in supporting these efforts. For example, it has been demonstrated that ChatGPT can match other models and is even superior to standard program repair techniques in debugging (Surameery & Shakor, 2023).

Thus, the following hypotheses address the research question of whether AI affects task performance. It is not a novelty that others, for example via outsourcing (e.g., Elmuti et al., 2010), or nowadays AI can support humans and take over work. Research has shown, for example, that when AI is correct, it can enhance human performance (Boyacı et al., 2023; Fügener et al., 2021). In this context, AI can also handle very complex tasks, such as assisting with programming (e.g., E. Chen et al., 2023; Y. Liu et al., 2023; Surameery & Shakor, 2023). Therefore, we assume that the code quality of a programming task will improve if GenAI is used to support it. This can be assessed in two ways. Firstly, through changes in the interaction with technology (Haleem et al., 2022), we assume that participants will perceive an increase in their own task performance (self-reported). Secondly, we aim to demonstrate this by measuring code quality using more objective metrics. Based on previous research on the use of ChatGPT in programming (e.g., Perkel, 2023) and in the evaluation of texts (Yi Chen et al., 2023), we expect that ChatGPT itself should be able to evaluate the task performance from participants' code snippets (AI-rated). Therefore, we hypothesize:

H1: Task performance increases in tasks with assistance of generative AI.

6.2.2.2 Meaningfulness

In addition to task performance, it is important to consider how people feel about their work concerning a programming use case. Nowadays, people aim to find meaning in their daily work (Cech, 2021; Jachimowicz & Weisman, 2022). Meaningfulness, or meaningful work, is defined as work that is worthwhile to oneself or personally significant (Allan, 2017). The benefits of meaningful work tasks can result in more motivation and, thus, higher performance of individuals, from which organizations can profit (Allan, 2017).

However, within the context of Information Systems (IS) research, the concept of meaningful work has often been relegated to describing results or emphasizing the significance of research findings. A structured literature review in the field of IS reveals a gap in the consideration of meaningfulness as a distinct construct. Despite the plethora of research in IS on well-being, only a limited number of studies have delved into the conceptualization and measurement of meaningfulness. Barkhi and Kao (2011) explore the psychological dimension of meaningfulness in the context of decision-making within Group Decision Support Systems (DSS). They find that users with higher levels of psychological meaningfulness make better decisions when they clearly understand the goal. Ke and Zhang's (2011) investigation into Open-source Software projects demonstrates that meaningfulness has a slightly negative influence on performance. The work of Kisekka and Goel (2022) introduces the concept of job meaningfulness as a factor influencing job performance, particularly during extreme events. Finally, Liu et al. (2017) see meaningful engagement in the gamification context as consisting of two elements, experiential and instrumental outcomes, thus also not directly measuring meaningfulness.

Therefore, it is essential to note that the term "meaningfulness" has not been consistently defined and utilized across various studies within the IS domain. The heterogeneity in the conceptualization of perceived meaningfulness underscores the need for a comprehensive and unified understanding of this construct to facilitate more coherent and comparable research outcomes.

Nonetheless, there have been ongoing discussions across disciplines about whether AI can increase or decrease perceived meaningfulness. AI, in general, can either replace human work, create new forms of collaboration, or amplify workers' skills (Bankins & Formosa, 2023). If AI automates human labor, the question arises whether humans lose the meaning in their work (Lysova et al., 2023) or if AI can relieve humans from repetitive or even dull tasks, thus increasing perceived meaningfulness (Bankins & Formosa, 2023). Especially with regard to the issue that people cannot always maintain the same passion for their work every day (Bredehorst et al., 2023), GenAI might either support individuals by reducing cognitive load or decrease perceived meaningfulness because the work feels less meaningful for them (Gielnik et al., 2015). As a result, the question remains about how the interaction changes when GenAI performs tasks that previously triggered creativity and added meaning to the job (Bankins & Formosa, 2023). We will aim to examine these questions in the current research.

Thus, the following hypotheses address the research question of whether AI affects the perceived meaningfulness of work concerning a programming use case. As concluded by Noy and Zhang (2023), the use of GenAI can impact job satisfaction. On the other hand, GenAI will automate human work, thus, meaning in their work might be lost (Lysova et al., 2023). In addition, programming is a rather creative task and not a repetitive one that can be automated to relieve humans (Bankins & Formosa, 2023). Although existing research does not provide a clear direction, we anticipate that the use of GenAI will generally have a negative effect on the perception of meaningfulness in programming. Therefore, we hypothesize:

H2: Perceived meaningfulness decreases in tasks with assistance of generative AI.

Finally, the next two hypotheses pertain to the mediation of our direct effects through perceived effort and task difficulty. It should logically follow that tasks that are easier to complete would also result in higher task performance. However, the concept of perceived meaningfulness adds complexity to the scenario. As technology can take over mundane task, it should feel more meaningful for individuals as they need to exert less effort to complete the same task (Bailey et al., 2019; Bankins & Formosa, 2023). However, research in psychology has indicated that the exertion of increased effort can sometimes make tasks feel more meaningful for individuals (Campbell et al., 2022; Gielnik et al., 2015; Inzlicht & Campbell, 2022; Mortimer, 2023). Additionally, research on human-AI collaboration suggests that humans prefer challenging tasks and do not delegate them to AI systems (Fügener et al., 2022). Thus, if AI can relieve humans from repetitive or even monotonous tasks (Bankins & Formosa, 2023), lower task difficulty and reduced effort invested may actually increase performance but decrease perceived meaningfulness as challenging tasks are reduced. Therefore, we hypothesize:

H3: Assistance of generative AI predicts higher performance by reducing task difficulty.

H4: Assistance of generative AI predicts lower perceived meaningfulness by reducing perceived effort.

6.3 Methodology

To obtain our data, we conducted an online experiment on Prolific Academic involving experienced Python programmers. Our primary goal was to examine (i) the performance of programmers and (ii) the perceived meaningfulness of their work when instructed to complete tasks with the assistance of AI. Online experiments are well-suited for manipulating specific conditions, enabling us to evaluate our measures of interest after each task. This approach is commonly employed in comparable contexts within the field of IS (e.g., Turel et al., 2008).

6.3.1 Research Design and Measurements

To assess the effect of GenAI on task performance and perceived meaningfulness, we instructed participants to undertake two basic programming tasks in Python, one with AI assistance and one without AI assistance (randomly assigned). We chose ChatGPT as the GenAI that assists our participants, as it is well-known by them. We hypothesized that engaging in everyday work tasks with AI, such as programming, would alleviate task difficulty, thereby enhancing task performance (Boyacı et al., 2023; Fügener et al., 2021). However, it also reduces the required effort, decreasing participants' perceived meaningfulness (Gielnik et al., 2015).

For higher levels of external validity, we created two groups with different types of typical programming questions and randomly assigned participants to one of them. They were either asked to write code from scratch (coding group) or instructed to find mistakes in existing code (debugging group). The tasks within each group were typical for learning programming, such as writing code for Fibonacci sequences or finding anagrams (see Figure 1). All tasks were developed in collaboration with a computer science professor and an experienced computer science student. Afterward, we piloted the tasks and the survey with five Ph.D. students in IS with substantial programming experience. Based on their feedback, we adjusted the wording of the tasks and survey questions to enhance the study's comprehensibility.

Imagine you are working together with an inexperienced colleague. Your boss asks you to write a program in Python to find anagrams (words or word sequences that are created by rearranging the letters of a word). The colleague offers to deliver a first draft, which you are supposed to correct and revise now:

```
# function to check if two strings are anagrams or not
def are_anagrams(s1, s2):

        # the strings are checked
        if(sorted(s1)== s2):
        return "The strings are anagrams."
        else:
        return False

# Testing the function
s1 ="listen"
s2 ="silent"
are_anagrams(s1, s2)
```

Your boss has explicitly recommended ChatGPT to you to ease your work. Therefore you ask ChatGPT for the errors in the code.

You get the following suggestion from ChatGPT:

Here are the identified errors in the code:

1. The sorted function is missing for s2. In order to compare sorted versions of both strings, sorted(s2) should be used.
2. If the two strings are anagrams the message "The strings are anagrams. is printed and if not False is returned.

Please make the necessary revisions to the code provided above and submit the updated code in the text box below. Please bear in mind that while ChatGPT strives for accuracy, it may not always be error-free. You can use spaces instead of tabs.

What errors did you identify and fix?

Figure 6-1 – Example Task in Debugging Group With ChatGPT

The experiment was structured as follows (see Figure 2): After participants agreed to a privacy policy, each participant was randomly assigned to one of the two groups: either coding from scratch or finding mistakes. Participants were asked to submit their code (completely new code for the coding group or improved code for the debugging group). If applicable, participants had to provide their prompts and the output of ChatGPT. Following each task, participants answered questions concerning their task performance (adapted from Goodman & Svyantek, 1999), perceived meaningfulness (adapted from May et al., 2004), perceived effort (adapted from Paas, 1992), perceived difficulty (on a 5-point Likert scale; adapted from Ribeiro and Yarnal, 2010), and the extent to which they utilized ChatGPT or other tools (e.g., Google or similar search engines) in the task (as a manipulation check). These measures were adapted from existing surveys to ensure validity and reliability. Finally, after completing the two tasks, participants answered demographic questions regarding their age and gender (see Table 1), as well as control variables related to their attitude toward AI, experience with coding, and coding enjoyment (see the Appendix for the constructs and items). Most variables were assessed on a 7-point Likert scale ranging from 1 = *Completely Disagree* to 7 = *Completely Agree*. All variables were integrated into an online survey tool and were marked as mandatory.

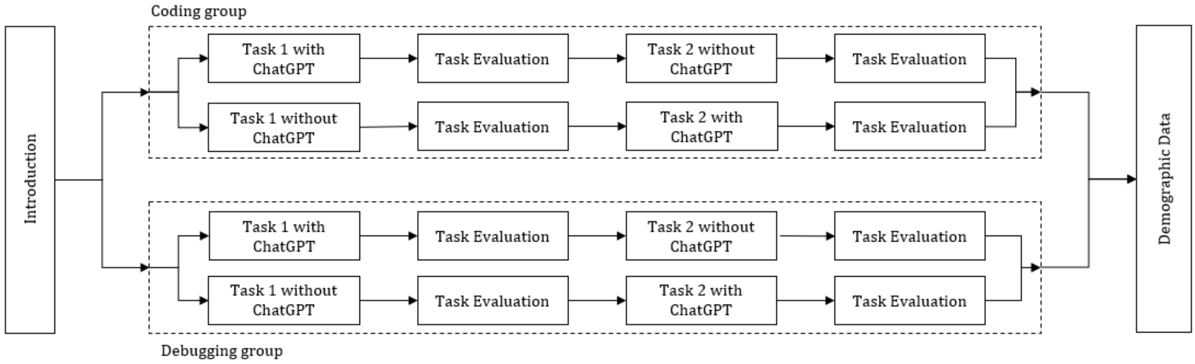

Figure 6-2 – Experiment Structure (Double Arrow Indicates Random Allocation)

In addition to assessing task performance based on the subjective responses from participants, we also evaluated task performance by reviewing the code and identifying errors in the task of error detection. Consequently, similar to Chen et al. (2023) for text quality, we requested ChatGPT to rate the quality of the code on a scale from 1 to 10, considering various aspects of code quality, such as whether the code fulfills the task, its efficiency, adherence to coding standards, and so on. After ChatGPT rated all 322 coding snippets, the authors randomly reviewed the code and ratings and generally concurred with the assigned scores.

6.3.2 Data Collection

After developing and testing our experiment, we administered it and collected data via Prolific Academic, allowing us to filter for our target group, which consisted of individuals fluent in English and experienced software developers in Python. We successfully collected the intended sample size of 160 participants. We excluded and replaced three participants who did not have access to ChatGPT, which resulted in a final sample size of 161. The participants were rewarded with about 10 €/h and completed the experiment in a mean time of 16 minutes. Most participants fell within the range of 18-34, and the gender distribution was balanced, as we had specified during the recruitment process on Prolific (see Table 1). Additionally, participants in our sample were relatively experienced coders (M = 5.26, SD = 1.12) who also enjoyed coding (M = 5.58, SD = 1.16). Furthermore, participants had favorable attitudes toward AI (M = 5.57, SD = 1.18).

Demography	Categories	Frequency	Demography	Categories	Frequency
Age	18-24	61 (38%)	Gender	Female	79 (49%)
	25-34	72 (45%)		Male	79 (49%)
	35-44	18 (11%)		Non-Binary	3 (2%)
	45-54	8 (5%)			
	<55	2 (1%)			

Table 6-1 – Demographic Data (Age and Gender) of the Participants

6.4 Results

In the following section, we will first present the results for the coding group, followed by the results for the debugging group. For both groups, we calculate Cronbach's alpha, composite reliability (CR), and average variance extracted (AVE) to confirm the internal consistency of our measures (Fornell & Larcker, 1981). Afterward, we will conduct paired *t*-tests to examine whether participants performed better and reported higher levels of perceived meaningfulness in the tasks with the assistance of ChatGPT. Finally, we will perform a mediation analysis to assess whether the relationship between the manipulation (task completion with and without ChatGPT's assistance) and task performance, as well as perceived meaningfulness, is mediated by perceived task difficulty and the effort invested in the task.

6.4.1 Coding Group

6.4.1.1 Reliability and Validity

Table 2 depicts the factor loadings for each variable, which are higher than the factor loadings on all other constructs, indicating convergent and discriminant validity. In addition, discriminant validity is verified, as the square root of AVEs (bold on the diagonal) is greater than all inter-construct correlations (Fornell & Larcker, 1981). All constructs' values for Cronbach's alpha, CR, and AVE were above the thresholds of 0.7, 0.7, and 0.5, respectively, thus confirming internal consistency (Hair Jr. et al., 2006, 2017). For this group, no significant correlation between task performance and perceived meaningfulness was found.

Construct	**Factor loadings**	**Cronbach's Alpha**	**CR**	**AVE**	1	2
1. Self-reported Performance	0.835-0.901	0.888	0.920	0.743	**0.862**	
2. Perceived Meaningfulness	0.834-0.925	0.964	0.959	0.800	-0.026	**0.893**

Table 6-2 – Reliability, Validity, and Correlation Matrix (Bold Numbers Are the Square Root of Aves; Significant Correlations (p < .05) Are Marked With *)

6.4.1.2 Hypotheses Test

We conducted paired *t*-tests between our manipulation (task completion with and without ChatGPT's assistance) and self-reported task performance and AI-rated task performance (H1), as well as perceived meaningfulness (H2). In line with H1, our results indicate that the use of ChatGPT predicted a higher self-reported task performance for participants (M = 6.25, SD = 0.87) compared to self-reported task performance without the assistance of ChatGPT (M = 6.09, SD = 1.01), $t(77)$ = 1.8, p < .050. To measure the effect size, we also calculate Cohen's d. Here, the effect size d = 0.20 indicates a small effect. Also, our results indicate that the use of ChatGPT predicted a higher AI-rated task performance for participants (M = 7.47, SD = 0.66) compared to AI-rated task performance without the assistance of ChatGPT (M = 6.47, SD = 2.14), $t(77)$ = 3.9, p < .001 with an effect size of d = 0.44, indicating a medium effect. Thus, H1 was supported by our data. In line with H2, our results indicate that the use of ChatGPT predicted a lower perceived meaningfulness for participants (M = 3.88, SD = 1.58) compared to perceived meaningfulness without the assistance of ChatGPT (M = 4.35, SD = 1.49), $t(77)$ = -3.7, p < .001 with an effect size d = 0.42, indicating a medium effect. Thus, H2 was also supported by our data.

Next, we conducted a structural model analysis based on our hypotheses H3 and H4 using a bootstrapping procedure involving 5,000 subsamples in SmartPLS 4 (Kushary et al., 2000) (see Figure 3 for the complete model). In particular, we estimated a mediation model from our manipulation (the assistance of ChatGPT or not) to our outcome variables, including self-reported and AI-rated task performance, as well as perceived meaningfulness, through the two mediators perceived task difficulty and effort. Our model demonstrates that the independent variables and the mediation account for a significant portion of the variance in the dependent variables (see R^2 for each construct). We assessed all hypotheses by examining the path coefficient and its significance using p-values.

First, we found a significant relationship between the use of ChatGPT and task difficulty (*b* = -0.630, *p* < .001), indicating that using ChatGPT resulted in decreased task difficulty for the participants. Similarly, they perceived their tasks as less effortful when they used ChatGPT (*b* = -0.786, *p* < .001).

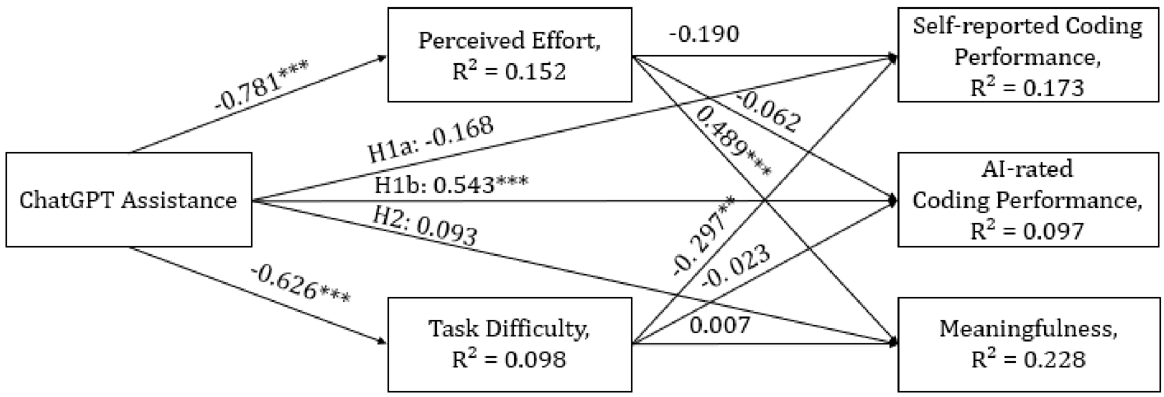

Figure 6-3 – Results of the Structural Model for the Coding Group (* p < .05, ** p < .01, *** p < .001)

Second, task difficulty was negatively related to self-reported task performance (*b* = -0.297, *p* < .010) but showed no significant connection with AI-rated task performance (*b* = -0.023, *p* > .050). Task difficulty was also not significantly related to perceived meaningfulness (*b* = 0.007, *p* > .050). On the other hand, perceived effort did not have a significant effect on self-reported task performance (*b* = -0.190, *p* > .050) or AI-rated task performance (*b* = -0.062, *p* > .050) but had a positive effect on perceived meaningfulness (*b* = 0.489, *p* < .001). As a result, the indirect effect between the use of ChatGPT and outcome variables through task difficulty was only significant for self-reported task performance (*b* = 0.186, *p* < .050). In contrast, the indirect effect between the use of ChatGPT and outcome variables through perceived effort was only significant for perceived meaningfulness (*b* = -0.382, *p* < .001). Thus, H3 was partially supported and H4 was supported by our data. Finally, when including the mediating variables, the direct effect between the use of ChatGPT and self-reported task performance became non-significant (*b* = -0.168, *p* > .050). Likewise, the direct effect between the use of ChatGPT and perceived meaningfulness was also non-significant (*b* = 0.093, *p* > .050). However, the direct effect between the use of ChatGPT and AI-rated task performance remained significant (*b* = 0.543, *p* < .001), as we did not identify a significant mediator for this relationship. Taken together, the effect of the use of ChatGPT on self-reported task performance was fully mediated via perceived task difficulty, and the effect of the use of ChatGPT on perceived meaningfulness was fully mediated via perceived effort.

6.4.2 Debugging Group

6.4.2.1 Reliability and Validity

Again, Table 3 shows the factor loadings for each variable, which are higher than the factor loadings on all other constructs, indicating convergent and discriminant validity. In addition, discriminant validity is verified as the square root of AVEs (bold on the diagonal) are greater than all inter-construct correlations (Fornell & Larcker, 1981). All constructs' values for Cronbach's alpha, CR, and AVE were above the thresholds of 0.7, 0.7, and 0.5, respectively, thus confirming internal consistency (Hair Jr. et al., 2006, 2017). Here, a positive and significant correlation between task performance and perceived meaningfulness can be found.

Construct	**Factor loadings**	**Cronbach's Alpha**	**CR**	**AVE**	1	2
1. Self-reported Performance	0.828-0.915	0.910	0.930	0.769	**0.877**	
2. Perceived Meaningfulness	0.864-0.916	0.965	0.976	0.785	0.202*	**0.886**

Table 6-3 – Reliability, Validity, and Correlation Matrix (Bold Numbers Are the Square Root of Aves; Significant Correlations (p < .05) Are Marked With *)

6.4.2.2 Hypotheses Test

In line with our procedure for the first group, we conducted paired t-tests between our manipulation (task completion with and without ChatGPT's assistance), self-reported task performance and AI-rated task performance (H1), and perceived meaningfulness (H2). In contrast to the first task, our results indicate that the assistance of ChatGPT is not significant regarding a higher self-reported task performance for the participants (M = 5.99, SD = 0.92) compared to self-reported task performance without the assistance of ChatGPT (M = 5.98, SD = 1.03), $t(82) = 0.1$, p > .050 with an effect size of d = 0.01, indicating a very small effect. However, our results indicate that the use of ChatGPT predicted a higher AI-rated task performance for participants (M = 5.20, SD = 2.72) compared to AI-rated task performance without the assistance of ChatGPT (M = 3.77, SD = 1.93), $t(82) = 3.6$, p < .001 with an effect size of d = 0.40, indicating a medium effect. Thus, H1 is only partially supported by our data. In contrast to the first group, H2 was not supported by our data, given that the assistance of ChatGPT did not indicate a higher perceived meaningfulness for the debugging group (M = 4.69, SD = 1.66) compared to perceived meaningfulness without the assistance of ChatGPT (M = 4.75, SD = 1.59), $t(82) = -0.7$, p > .050 with an effect size of d = 0.08, indicating a very small effect.

Again, we conducted a structural model analysis based on our hypotheses H3 and H4 (see Figure 4 for the complete model). Our model demonstrates that the independent variables and the mediation account for a portion of the variance in the dependent variables (see R^2 for each construct). We assessed all hypotheses by examining the path coefficient and its significance using p-values.

First, we found a significant relationship between the use of ChatGPT and task difficulty (b = -0.293, p < .050) again, indicating that using ChatGPT resulted in decreased task difficulty for the participants. Similarly, participants perceived their tasks as less effortful when they used ChatGPT (b = -0.301, p < .050).

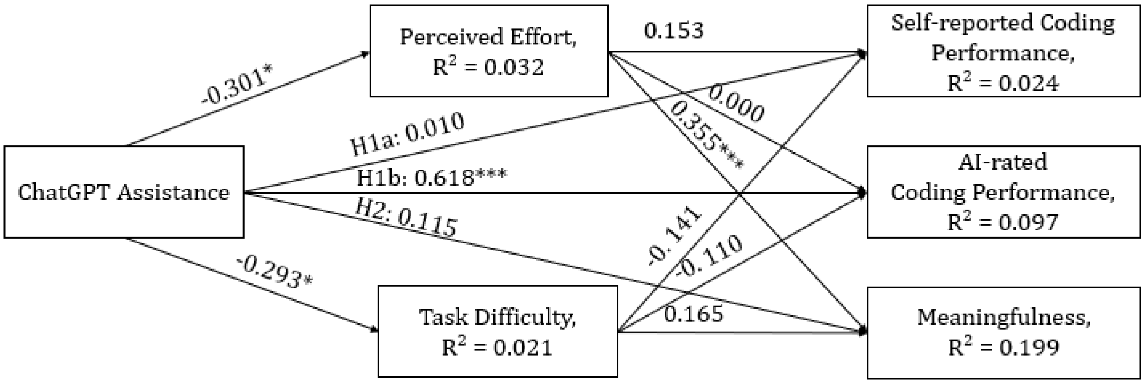

Figure 6-4 – Results of the Structural Model for the Debugging Group (* p < .05, ** p < .01, *** p < .001)

Second, the task difficulty had no significant effect on either the self-reported task performance (b = -0.141, p > .050), AI-rated task performance (b = -0.110, p > .050), or perceived meaningfulness (b = 0.165, p > .050). Thus, the indirect effect of ChatGPT assistance via task difficulty to self-reported task performance is not significant (b = 0.041, p > .050). Also, the perceived effort did not have a significant effect on self-reported task performance (b = 0.153, p > .050) or AI-rated task performance (b = 0.000, p > .050) but had a positive significant effect on perceived meaningfulness (b = 0.355, p < .001). However, the indirect effect of ChatGPT assistance via perceived effort to perceived meaningfulness was not significant (b = -0.107, p = .099). Since the effect between ChatGPT assistance and perceived effort and perceived effort and perceived meaningfulness are quite small, the indirect effects between the use of ChatGPT and the outcome variables through the mediators were non-significant. Thus, H3 was not supported and H4 was partially supported by our data. Finally, when including the mediating variables, the direct effect between the use of ChatGPT and self-reported task performance was still non-significant (b = 0.010, p > .050). Likewise, the direct effect between the use of ChatGPT and perceived meaningfulness was also not significant (b = 0.115, p > .050). However, the direct effect between the use of ChatGPT and AI-rated task performance remained significant (b = 0.618, p < .001). Taken together, our results for the second task were descriptively similar to those in the first task. However, we did not observe as many significant effects, suggesting that contextual factors between the two different tasks may either amplify or mitigate the effects we observed.

In summary, we can support H1 with the coding task and partially with the debugging task. We can accept H2 for the coding, but not the debugging task. H3 is partially supported by the coding, but not the debugging task and finally, we can accept H4 for coding and partially for the debugging task.

6.5 Discussion

AI has become an integral part of everyday life (Berente et al., 2021; Brynjolfsson & Mitchell, 2017; Russell & Norvig, 2021), especially with the emergence of GenAI tools such as ChatGPT (Dwivedi et al., 2023; Teubner et al., 2023). While the enhancement of decision-making and task performance through human-AI collaboration is well-documented (e.g., Boyacı et al., 2023; Fügener et al., 2021), the psychological well-being and perceptions of its users also can have an influence on the adoption of such technologies. Therefore, it is important to understand what influences the use of GenAI beyond task performance. Surprisingly, limited attention has been directed towards psychological constructs such as perceived meaningfulness in the context of AI. Initial studies, like those by Noy and Zhang (2023), suggest a positive correlation between

ChatGPT use and job satisfaction, which we intend to delve deeper into. After all, higher perceived meaningfulness can lead to greater motivation and, therefore, higher organizational performance (Allan, 2017). In line with this, our primary goal of this experiment was to examine the evaluation of coders' task performance and their perception of meaningfulness when engaged in tasks with AI support. Such insights can pave the way for more holistic strategies in AI adoption, ensuring that technological advancements are harmoniously aligned with human well-being and purpose. To address our research question and test the subsequent hypotheses, we divided our participants into two groups with two different assignments. This allows us to test our research questions with two different scenarios, reflecting ChatGPT's multifaceted programming potential. Thus, the coding group had to solve coding tasks, and the debugging group had to perform error detection tasks. Both groups had to fulfill one task with and the other without ChatGPT assistance, which served as our primary manipulation in this experiment.

The first part of our research question centered around the task performance augmentation with AI assistance. Task performance was measured in two ways: self-reported by the participants of the experiment and AI-rated with ChatGPT. Consistent with previous studies on human-AI collaboration (Boyacı et al., 2023; Fügener et al., 2021) we found that participants performed higher in their tasks when they had assistance from ChatGPT. This was shown in both groups through significant *t*-tests between the tasks solved with ChatGPT and without its assistance. In addition, in our structural model, a direct effect was particularly strong for the more objective AI-rated task performance, but we also found an indirect effect via task difficulty for self-reported task performance in the coding group. Furthermore, participants reported reduced effort and task complexity when assisted by ChatGPT. These observations underscore the potential of GenAI in simplifying tasks. Thus, we conclude that the assistance of ChatGPT improves the performance in complex activities such as coding.

Regarding the second part of our research question on how AI assistance impacts the sense of perceived meaningfulness in people's work, we identified significant differences with *t*-tests only for the use of ChatGPT in the coding group, indicating that participants reported lower levels of perceived meaningfulness when they used ChatGPT; the *t*-test in the debugging group, on the other hand, was not significant. However, the mediating effect through perceived effort emerged as significant for both groups, suggesting that for employees to rate a task as meaningful, they also need to feel that they have invested effort in it. This is intriguing because task performance is only mediated by task difficulty. Importantly, our findings are in line with emerging research in psychology that has indicated that increased effort can sometimes make tasks feel more meaningful for individuals (Campbell et al., 2022; Gielnik et al., 2015; Inzlicht & Campbell, 2022; Mortimer, 2023). Thus, our findings suggest that when people perceive tasks with ChatGPT assistance as less effortful, the meaningfulness decreases accordingly.

It is noteworthy that we also found nuanced differences between the two programming tasks. While not conclusive, the directionality of the relationship between perceived effort and self-reported task performance varied between the groups. For instance, a negative relationship was found in the coding group, whereas in the debugging group, a positive one was observed. Additionally, the influence of task difficulty on self-reported task performance in the second group is no longer significant. This may be attributed to the perception that the task is easier, as the ChatGPT support has a slightly less pronounced impact on the task difficulty and perceived errors. Consequently, the support of the GenAI no longer makes as significant a difference for the participants. Alternatively, it is plausible that coding is perceived as more creative, and thus, the use of GenAI has a greater influence than in debugging, in particular error-finding, which might be more repetitive and mundane.

6.5.1 Contributions

With our experiment, we can offer various theoretical contributions. As technology exceeds its traditional function of automating routine tasks and enters areas of creativity, it is crucial to assess the relationship between AI and human work. First, we align with the existing literature on human-AI collaboration and show that the use of AI results in higher task performance (e.g., Boyacı et al., 2023; Fügener et al., 2021). This is especially relevant for the specific coding tasks that have been investigated to a lesser extent in the past, as previous literature has mainly focused on AI for decision-making. We can, therefore, extent existing human-AI research with showing that this is also relevant for GenAI in coding tasks thus generalize on the one hand that GenAI leads to higher task performance and, on the other hand, that AI has a positive influence on these coding tasks. Second, our study aims to bridge the gap between the IS literature and meaningful work. Our paper addresses this void by systematically examining the role of perceived meaningfulness, particularly in the context of programming tasks. Thus, we also assess the psychological outcomes of how GenAI changes the way we work combining IS research with known phenomena from psychology (see Bailey et al., 2019). This is especially relevant as Gen Z, for example, rate meaningfulness as one of the main factors to take a new job (McKinsey, 2022). Depending on the task, we observed a decrease in perceived meaningfulness with ChatGPT assistance. It is possible that AI reduces the effort we have to invest in our work, leading to a perception of our work as less meaningful. Since these results are mixed, we can only provide an initial assessment of the influence of GenAI on perceived meaningfulness. Third, in the past, AI has been less capable of handling highly creative tasks (Dwivedi et al., 2023). This is now changing with GenAI. It was previously assumed that creative tasks, in particular, lead to a higher level of perceived meaningfulness, which is also demonstrated here to some extent. Thus, our research indicates that the focus should be on replacing or collaboratively solving relatively easy, boring, or repetitive tasks with (generative) AI, while leaving complex or creative tasks to humans that allow them to find meaning in their work (Cech, 2021; Jachimowicz & Weisman, 2022). Perhaps even new areas of work that are complex and particularly suitable for humans to take responsibility in the future will develop.

In addition to theoretical contributions, we also provide several practical ones. Firstly, companies should ensure that employees' perceived meaningfulness remains intact when working with GenAI. After all, perceived meaningfulness leads to more motivation, which, in turn, results in higher company performance (Allan, 2017). To achieve meaningfulness, it is important to consider which tasks can be performed collaboratively with AI, such as replacing routine tasks while jointly working on complex tasks or letting humans perform creative tasks. Additionally, the literature on job crafting offers models that can provide valuable guidance. Secondly, companies should allow and promote the use of GenAI within the company due to the increase in task performance caused by human-AI interaction. This is because higher individual performance contributes to higher company performance. Finally, while we recommend that employees be encouraged to use AI, companies should consider and evaluate which tools are most beneficial. ChatGPT can be utilized for texts or coding, but more specialized AI systems, such as GitHub Copilot for coding or DALL-E 3 for generating images, exist. In this way, GenAI can potentially reduce cognitive load and, therefore, create benefits for employees using such technologies. However, questions about the privacy and security of data entered into GenAI tools remain, as well as concerns about extending existing models with the company's own data.

6.5.2 Limitations and future research

Like any research, our study is not without limitations. First, our sample size is quite young, although software engineers are in general younger this is a limitation where future research might look into the effect on the effect on older people. While we utilized two different programming assignments in our experiment, they were still very specific. Thus, future research should aim to replicate our findings with additional (programming) tasks and different GenAI tools. Tools like Microsoft 365 Copilot might be interesting as it can change the entire scope of work, even the administrative area. In addition, we suggest exploring contextual effects that may either enhance or diminish the strength of our observed effects. Furthermore, although ChatGPT's responses were artificially predetermined in the debugging group, enhancing comparability might not accurately represent the interaction and, therefore, could impact the perceived meaningfulness. However, in the coding group, participants were required to interact with ChatGPT in real time during the coding task. The interaction, however, could only be assessed through control questions on the prompts and the output of ChatGPT. Finally, a high level of perceived meaningfulness might be attributed to the novelty of the situation. Although we made sure that the participants had experience with ChatGPT the novelty may diminish influencing the perceived meaningfulness—similar to experiences on outsourcing—if interactions with GenAI become routine. For instance, the literature on passion demonstrates significant fluctuations throughout daily work (Bredehorst et al., 2023). Therefore, future studies could potentially benefit from a later time of measurement and should be duly conducted.

6.6 Appendix

Table A shows the constructs, items, and control variables measured during the experiment:

Construct	Item Description	Source
Task Performance (self-reported)	I achieved the objectives of the task. I met criteria for performance. I fulfilled all the requirements of the task. I performed well in the overall task by carrying out tasks as expected.	Goodman and Svyantek (1999)
Perceived Meaningfulness	This task was very important to me. Completing this task was personally meaningful to me. The task I just did was worthwhile. This task was significant to me. The work I did on this task was meaningful to me. I feel that the work I did on this task was valuable.	May et al. (2004)
Perceived Effort	Please choose the level of effort that applied to you during this task.	Paas (1992)
Task Difficulty	How did you perceive the difficulty of this task?	Ribeiro and Yarnal (2010)
Attitude toward AI	How would you assess your attitude towards ChatGPT (or similar systems)?	-
Experience with Coding	How would you rate your coding experience (e.g., writing code for similar tasks like in our study)?	-

Construct	Item Description	Source
Coding Enjoyment	How enjoyable do you find coding (e.g., writing code for similar tasks like in our study)?	-

Table 6-4 – Constructs, Control Variables, and Their Items Measured in the Experiment

6.7 Acknowledgement

This research and development project is/was funded by the German Federal Ministry of Education and Research (BMBF) within the “Innovations for Tomorrow’s Production, Services, and Work” Program (funding number 02L19C150) and implemented by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication.

7 Paper E: The Influence of Effort on the Perceived Value of Generative AI: A Study of the IKEA Effect

Title

The Influence of Effort on the Perceived Value of Generative AI: A Study of the IKEA Effect

Authors

Mehler, Maren F., Ellenrieder, Sara and Buxmann, Peter

Publication Outlet

Proceedings of the Thirty-Second European Conference on Information Systems (ECIS 2024), Paphos, Cyprus

Abstract

While the use of Generative Artificial Intelligence (GenAI) aims to automate human tasks, psychology research shows how crucial human effort is for the appreciation of the final results. The so-called "IKEA effect" refers to the increased valuation individuals attribute to self-created products. However, the potential implications of this effect for GenAI have remained unexplored. This study delves into the presence of the IKEA effect in GenAI, specifically focusing on image creation. Through an online experiment involving 174 participants in Germany, we observed that participants valued images higher if more human effort was invested during collaborative co-creation with GenAI. Our findings indicate a significant presence of the IKEA effect, although existing GenAI research primarily focuses on the automation of processes. This discovery emphasizes the importance of understanding user psychology and also offers valuable insights for designing and leveraging GenAI applications.

Keywords

IKEA Effect, Generative Artificial Intelligence, Online Experiment, Human-AI Collaboration

7.1 Introduction

Cognitive biases have been shown to influence decision-making in various instances (e.g., Hristov et al., 2022; Ni et al., 2019), including the IKEA effect, which states that people tend to value self-assembled physical objects more than identical objects that were assembled by others (Norton et al., 2012). Attributing a higher perceived value to objects into which personal effort has been poured during development or assembling already significantly impacts how companies shape customer experiences today (Franke et al., 2010). Adidas, for instance, enables their customers to customize clothing through their online platform (Adidas, 2023). Similar design choices for customers are offered by the popular backpack brand Fjällräven (Fjällräven, 2023)(Fjällräven, 2023), and Build-A-Bear provides customers with the chance to design stuffed animals online and even participate in the stuffing process at their physical stores (Build-A-Bear, 2023). As a result, customers are evolving into active co-creators rather than merely passive consumers of products (Mochon et al., 2012).

Advancements in data availability, training techniques, and scale of model parameters have made Generative Artificial Intelligence (GenAI) models more versatile, capable, and applicable to a wide range of tasks and domains such as text or image creation (Dwivedi et al., 2023). Recent industry reports show that modern GenAI promises to automate tasks that currently consume 60 to 70 percent of employees' daily work hours. This heightened potential for technical automation primarily stems from the advancements in GenAI's proficiency in comprehending natural language—a critical requirement for tasks constituting 25 percent of the overall work hours (McKinsey, 2023). The automation goals pursued with the use of intelligent technologies are a paradox with the trend of allowing customers to participate in the product development process.

While the primary goals of the deployment of (generative) AI are the automation of tasks and minimization of human effort to successfully accomplish tasks (Berente et al., 2021; Brynjolfsson & Mitchell, 2017; Russell & Norvig, 2021), challenges for the success of this technology arise as psychology and behavioral research has already taught us that "labour leads to love" (Norton et al., 2012). While many studies have proven the IKEA effect for physical objects such as origami, food, or furniture (e.g., Dohle et al., 2014; Ling et al., 2020; Mochon et al., 2012; Norton et al., 2012), we see a clear lack of research for non-physical objects or content in general. Thus, we do not yet understand to what extent the IKEA effect also affects the perceived value of non-physical goods. At the same time, the research stream of human-AI collaboration investigates how humans and intelligent machines can work together synergistically to enhance problem-solving and decision-making processes (e.g., Abel-Karim et al., 2020; Boyacı et al., 2023; Fügener et al., 2022; Jussupow et al., 2021). However, several unique challenges and characteristics make it uncertain whether and how the IKEA effect could manifest. In previous research on the IKEA effect, humans could always estimate what activity and skill was needed to assemble or create goods such as furniture or food. Missing transparency (often known as black box behavior) in the AI algorithms increases the difficulty of comprehending how the AI makes decisions or creates results (Bauer, Hinz, van der Aalst, et al., 2021; Dwivedi et al., 2023). It is also unclear which influence humans have on the results through prompting and how the AI processes human inputs; only the results can be evaluated (Dwivedi et al., 2023; OpenAI, 2023b).

In this context, it is crucial to understand the impact that the takeover of human tasks by AI will have on the perceived value of the final results. Regarding the IKEA effect, it can be assumed that the perceived value of AI-generated content and users' behavioral intention to use GenAI tools increase, especially when they have the opportunity to put effort into the collaboration. Thus, we seek to answer the following research questions: *(1) Does human effort invested in collaboration with generative AI promote overvaluation of AI-generated solutions, and (2) does this heightened perceived value also increase the behavioral intention to use GenAI technology?*

To answer these research questions, we conducted an online experiment with 174 people who interacted with GenAI tools to perform work tasks with a high vs low level of effort in the collaborative creation of content. We hereby assess the impact that effort poured into the collaboration with the AI has on the perceived value of the generated content and the behavioral intention to use the technology. We contribute to research by examining how humans and AI should collaborate in the future to value the content created through collaboration. The results of our online experiment reveal that the IKEA effect is prevalent in the collaborative creation of content by humans and GenAI. Humans tend to overvalue AI-generated content if effort is invested into collaboration with the technology. Future research can build on this study to derive design guidelines for GenAI tools and strategies for human-AI collaboration that ensure users value results.

7.2 Theoretical Background

The following section outlines the unique characteristics of GenAI. It then provides an overview of cognitive biases, focusing on the IKEA effect and its impact on the perceived value of self-created objects so we can hypothesize the impact of the IKEA effect on human-AI collaboration afterward.

7.2.1 Artificial Intelligence

Berente et al. (2021) describe AI as the frontier of emerging technologies that is focused on human intelligence for complex decision-making. There are several subcategories of AI, such as machine learning (ML), which are learning algorithms that recognize patterns in data to make decisions or predictions (Brynjolfsson & Mitchell, 2017; Mitchell, 1997; Russell & Norvig, 2021). Or, more recently, the advances in GenAI such as ChatGPT, Google's Bard, or other Large Language Models (LLMs), which differ from ML in creating new data from recognized patterns and not only analyze existing data (Dwivedi et al., 2023; Teubner et al., 2023). However, these language models are not the only novelty in this area. AI image creation tools like Stable Diffusion, Dall-E 2, or Midjourney also prove the capabilities of AI as generated images are very realistic and appealing (Göring et al., 2023). The use of ChatGPT is currently being discussed in radiology (Rao et al., 2023), for cybersecurity (Prasad et al., 2023), and in a wide range of business areas. Although such systems can already facilitate work today, especially in the future (Dwivedi et al., 2023), caution must be taken that jobs are not destroyed and that users accept the outcomes of AI.

AI is characterized, in particular, by the three properties autonomous, inscrutable, and self-learning (Berente et al., 2021). In addition, the potential for bias in training data, the ownership of the training data as well as the output, and the potential of wrong outputs are concerns in GenAI (Lund et al., 2023). In particular, the "black box" characteristic of AI makes it difficult for humans to follow the decision of AI and thus understand and adapt the given output (Bauer, Hinz, Aalst, et al., 2021). This characteristic remains applicable in GenAI as well (Dwivedi et al., 2023). The research stream on human-AI collaboration investigated the influence of machines and humans working together (e.g., Boyacı et al., 2023; Fügener et al., 2022). It was shown that humans want to decide rationally but often cannot, e.g., due to a misjudgment of the task's difficulty (Fügener et al., 2022). The research also shows that people are reluctant to perform simple tasks and enjoy demanding tasks themselves, which makes delegation to AI problematic (Fügener et al., 2022) . In addition, it was demonstrated that when the results of decisions differ between humans and AI, people have different approaches to dealing with it. For example, experienced physicians tend to ignore AI advice, while novice physicians question their own decisions and are less satisfied with the AI system (Jussupow et al., 2021). It is also essential that human knowledge is not lost in the interaction between humans and AI but should be actively incorporated into decision-making (Fügener et al., 2021). Hence, there should be a focus on developing systems that allow humans and AI to work together rather than AI exclusively reaching final decisions (Abel-Karim et al., 2020).

7.2.2 Cognitive Bias

Humans often deviate from optimal decisions since always thinking and acting rationally is impossible. This is due to several factors: People usually cannot process the entire amount of information or do not have it available; when processing the information, incorrect conclusions can be drawn, for example, due to interpretation influenced by emotions or existing prejudices; or heuristics are used when making quick decisions (Kahneman, 2012). These influences on rational decisions are called cognitive biases. They can be grouped into four categories: First, *information and perception biases* can occur, such as the anchoring bias, which refers to the tendency to rely on the first information that comes to mind when making decisions (Kahneman, 2012). Second, *decision biases* exist, such as the overconfidence bias, which is the tendency to overestimate one's own judgment. It leads to a misplaced sense of certainty in decision-making, making decisions riskier (Odean, 1998). In addition, *social and emotional biases* are prevalent, such as the self-serving bias, which is the tendency to make decisions that benefit one's own interests (Miller & Ross, 1975). Lastly, *technology and change biases* affect decisions, such as the status quo bias, which describes the tendency to prefer the current state of affairs and resist change. This can lead to rejecting new technologies, ideas, or practices even if they offer improvements (H.-W. Kim & Kankanhalli, 2009).

Especially in IS research, there is a lot of literature on cognitive biases in decision-making (e.g., Hristov et al., 2022; Ni et al., 2019) or biases in the adoption of emerging technologies (e.g., Balakrishnan et al., 2021; Frank et al., 2023; Piehlmaier, 2022). For example, Phillips-Wren et al. (2019) show that overconfidence bias is an inhibitor in adopting and using decision aids. However, Piehlmaier (2022) finds that overconfident investors are more likely to use Robo advisors. In consumer adoption, another exemplary bias, the negativity bias, is evident (Frank et al., 2023) or individuals do not always choose the best algorithm for their decisions (Dietvorst & Bharti, 2020). In addition, Kim and Kankanhalli (2009) illustrate that user resistance to information systems is related to status quo bias. Similarly, Balakrishnan et al. (2021) show status quo bias as a factor in accepting AI-powered voice assistance. Furthermore, a focus on organizations can often be found. In this regard, Ni et al. (2019) show that anchoring bias occurs in a corporate context when making decisions using a BI system. More generally, Hristov et al. (2022) show what possible cognitive biases occur within decisions about performance management systems. Once biases are known, they can be prevented or exploited to force rational or reasonable decisions.

7.2.3 Understanding the IKEA Effect

One bias that can be best categorized as a *decision bias* and is widely studied to explore the impact of user participation on the success of products or services is the so-called IKEA effect. The IKEA effect states that people value objects higher if they assemble or create them on their own (Norton et al., 2012). Norton et al. (2012) named this phenomenon after the Swedish manufacturer whose products are particularly likely to involve a high level of assembly effort. Higher effort means a person must invest more work in a task (see Marsh et al., 2022). While the IKEA effect is unsurprising for some products such as art, studies have demonstrated the IKEA effect for a variety of physical objects such as food, Lego, clothing, or simple IKEA cardboard boxes (e.g., Dohle et al., 2014; Ling et al., 2020; Mochon et al., 2012; Norton et al., 2012; Radtke et al., 2019). In fact, the value of self-created objects is estimated even higher than the value of objects created by experts (Norton et al., 2012) and the IKEA effect also remains prevalent in collaborative creation (Marsh et al., 2022). In addition, the effort that goes into the creation of the objects does not have to be associated with fun for consumers to overvalue their creations (Mochon et al., 2012; Norton et al., 2012). However, successfully completing a task is necessary for the IKEA effect to emerge (Norton et al., 2012).

Evidence that "labour leads to love" (Norton et al., 2012) can also be derived from the success of past product launches. When instant cake mixes were introduced to minimize the manual labor required for baking, the initial success failed as the preparation of a cake was now perceived as too simple. The recipes were changed so that the addition of eggs was required. Although other influences may have been at work here, the subsequent adoption success is often attributed to the IKEA effect, as the customer puts additional work into the product (Norton et al., 2012; Shapiro, 2004). Today, multiple companies take advantage of the IKEA effect to shape customer experience by providing opportunities to customize and create products online through configurators, toolkits and choice menus (Franke et al., 2010). Therefore, customers can increasingly be described as co-creators instead of passive recipients of goods (Mochon et al., 2012). Providing ideas, thoughts, feelings, and, most importantly, actions to participate in the co-creation process leads to a higher perceived value, which has been shown to increase the willingness to pay (WTP) for these products (Mochon et al., 2012; Norton et al., 2012).

Although the bias has been successfully demonstrated in many studies (e.g., Dohle et al., 2014; Ling et al., 2020; Marsh et al., 2022; Norton et al., 2012), uncertainty exists regarding the actual cause. Psychology and behavioral research suggest three primary mechanisms that cause the IKEA effect: (a) signal of competence, (b) effort justification, and (c) ownership (Marsh et al., 2018). First, self-created objects can express the competence of the creator and can be used as a trophy to show off, increasing the perceived value of the object for the creator (Mochon et al., 2012). Second, within the concept of effort justification, creations reflect the investment of effort rather than signal competence (Norton et al., 2012). Thus, the increased value of a created object may reflect the effort invested. Lastly, the creation of an object can lead to ownership claims in some scenarios or promote the sense of ownership of the person who created the object (Kanngiesser et al., 2010; Marsh et al., 2022). Since people tend to place a higher value on their personal belongings compared to equivalent items that they do not own, this can further enhance the IKEA effect (Kahneman et al., 1990). Nevertheless, there is disagreement about the underlying mechanisms and it remains difficult to adequately explain the IKEA effect (Marsh et al., 2018).

A related research stream can also be found in the area of mass customization (MC), where customers are provided with easy-to-use configurators to design products themselves online, which are then produced by the manufacturer (Franke et al., 2010; Ling et al., 2020). Research on MC toolkits, however, is based on two assumptions: firstly, that *preference fit* is the essential benefit for customers, while *design effort* represents costs for the customer and should therefore be kept to a minimum (Franke et al., 2010; Randall et al., 2007). These two goals are consequently in conflict with each other, and MC toolkits should balance them optimally to gather enough information from customers to customize products according to their preferences while keeping the effort low. However, the IKEA effect contradicts these assumptions, suggesting that regardless of personal preferences, effort alone leads to a perceived increase in value (Norton et al., 2012). The so-called "I designed it myself" effect has also been shown in the context of MC toolkits, indicating that a feeling of accomplishment arises from creating a self-designed product, challenging the existing concept (Franke et al., 2010). While it is important in research on virtual co-creation tools for customers to contribute their own preferences in the development process, the IKEA Effect also occurs independently of this, even with simple products like IKEA cardboard boxes (Norton et al., 2012).

Furthermore, to the best of our knowledge, the IKEA effect has only been studied for physical products (e.g., Dohle et al., 2014; Norton et al., 2012) even though labor can also be invested into the creation of non-physical products or general content. In addition, the IKEA effect has been explored by comparing if people value their own creations more than similar objects created by others. However, the creation of physical as well as non-physical objects and content can also be performed by machines. Thus, we are interested if this will affect the perceived value of the results and whether the human effort involved in creating content or objects through extensive prompting will continue to be of great importance in the future.

7.3 Hypothesizing the IKEA Effect in Human-AI Collaboration

Our study combines insights from psychology on the IKEA effect (e.g., Norton et al., 2012; Radtke et al., 2019) with the research fields of human-AI collaboration in the context of GenAI and suggests that contrary to current efforts to replace human activities by GenAI (Dwivedi et al., 2023; Teubner et al., 2023), humans need to contribute effort to collaborative development in order to fully value results (e.g., Marsh et al., 2022). Although the black box nature of GenAI makes it difficult for humans to assess how much of an impact they have on the final outcome and how the AI works (Berente et al., 2021; Dwivedi et al., 2023; Lund et al., 2023), we hypothesize that the IKEA effect will occur, increasing WTP as shown in other IKEA effect studies on physical products (e.g., T. Liu et al., 2023; Marsh et al., 2022; Norton et al., 2012). Norton et al. (2012) showed, for example, that the WTP for origami symbols was significantly higher for self-created symbols compared to ones created by experts. In addition to studies showing that WTP increases for self-made objects, Walasek et al. (2017) further demonstrated that the price at which one would sell self-made products (in this case, assembling various science kits) is also significantly higher compared to the identical objects assembled by someone else. The fact that effort can increase WTP has already been shown for various physical products (e.g., Marsh et al., 2022; Walasek et al., 2017) and can also apply to content collaboratively created with GenAI if a lot of personal effort has been put into the collaboration. Moreover, a debate on WTP for GenAI arose with the introduction of ChatGPT (Capgemini, 2023), which we hope to contribute to. Thus, we propose:

H1: Willingness to pay for AI-generated solutions increases when effort is put into collaborating with generative AI.

Other studies have also shown that people often find it difficult to describe why they prefer a self-created product over the same product created by someone else and that this is often described as a personal feeling of *liking* or *appreciating* the self-created product or solution more (Dohle et al., 2014; T. Liu et al., 2023; Norton et al., 2012). Since we expect personal feelings to arise even with solutions that can be created by GenAI, such as images or texts, we propose:

H2: Appreciation for AI-generated solutions increases when effort is put into collaborating with generative AI.

Although research around the IKEA effect measures the perceived value of self-created products through WTP or abstract constructs such as *Liking*, *Niceness*, or *Appreciation* (e.g., Dohle et al., 2016; T. Liu et al., 2023; Norton et al., 2012), marketing and IS research proposes four dimensions that are positively associated with the overall *perceived value,* namely quality value, emotional value, value-for-money and social value (Turel et al., 2007). To measure whether the perceived value of AI-generated solutions increases as a result of investing effort, we build on the hypotheses of Turel et al. (2007). Within the realm of service-oriented marketing, research has demonstrated that superior quality assessments contribute to greater overall value (Baker et al., 2002; Brady & Robertson, 1999). Second, emotional components such as joy can enhance the formation of an overall value that individuals perceive, and third, users are price sensitive when evaluating the trade-off value-for-money (Turel et al., 2007). Lastly, individuals have the potential to improve their self-concept through the utilization of modern technologies, such as GenAI. This is because such technology can be perceived as cutting-edge and innovative, thereby signaling the user's affiliation with a specific social class (Schewe & Dillon, 1978). Thus, social value is also positively associated with the overall perceived value (Turel et al., 2007). Consistent with the IKEA effect (Norton et al., 2012), we propose that overvaluation of AI-generated solutions occurs when humans invest effort in collaborative development. This results in a higher perceived value (***H3***) which is driven by quality value (***H3.a***), emotional value (***H3.b.***), value-for-money ***(H3.c***), and social value (***H3.d***):

H3: Perceived value of AI-generated solutions increases when effort is put into collaborating with generative AI.

H3.a: Quality value of AI-generated solutions increases when effort is put into collaborating with generative AI.

H3.b: Emotional value of AI-generated solutions increases when effort is put into collaborating with generative AI.

H3.c: Value-for-money of AI-generated solutions increases when effort is put into collaborating with generative AI.

H3.d: Social value of AI-generated solutions increases when effort is put into collaborating with generative AI.

Based on this, Turel et al. (2007) demonstrated—without reference to the IKEA effect—that higher perceived value is a key determinant of behavioral usage intentions. They conclude that only if the benefits of a technology are clear, it will be utilized. Similarly, Kamtarin (2012) demonstrated that perceived value as a positive effect on online purchase intention. As we expect that people perceive a higher value of content generated through the interaction with GenAI, they will learn about the benefits and thus have a higher intention to reuse it. Therefore, we hypothesize that with a higher perceived value of AI-generated content, behavioral usage intentions for generative AI will increase:

H4: Behavioral intention to use generative AI increases when effort is put into collaborating with the technology.

7.4 Methodology

To answer our research question of whether the IKEA effect increases the perceived value of content created in collaboration with GenAI tools, such as ChatGPT or Stable Diffusion, we conducted an experiment, which is often utilized in IS research on human-AI collaboration (e.g., Fügener et al., 2022) as well as IKEA studies (e.g., Norton et al., 2012). Our goal is to investigate whether the IKEA effect exists for GenAI, meaning that people overvalue AI-generated content when they have put their own effort into a collaborative development process. We defined our target group as (working) individuals in Germany who might use GenAI now or in the future for their daily life/ work.

7.4.1 Research Design and Measures

For the experiment, we developed two tasks to collaboratively generate content with AI: one task where participants had to generate a text about a company mission statement for a drone start-up in the healthcare sector using ChatGPT and one task where the participants had to generate an image for advertising sustainably packaging apples using Stable Diffusion. GenAI can be used to create different types of content such as audio, code, images, text, simulations and videos. IKEA studies have already shown that the IKEA effect can occur with physical objects when creating pictures (Raghoebar et al., 2017). We decided to choose images and text for the overall design of the experiment because, firstly, both are content that humans can generate without the addition of technology and it is easier to assess the effort and value of the end result without specialist knowledge. Secondly, there are powerful solutions for generating both types of content on the market that are already widely used in everyday professional life. We have explicitly excluded code here, as functionality and lean implementation are the main focus when generating code and individual preferences are often given less importance. For both tasks we created an IKEA group, which allowed participants to put effort into the collaboration with the AI, and a control group respectively, which was confronted with the final results of the GenAI tools immediately after reading the problem definition. Participants were randomly assigned to the IKEA group for one task and then assigned to the control group for the other task. Therefore, each participant solved one task (image or text) for the IKEA group and the other task for the control group. Task I (text) and task II (image) and thus the sequence of participation in the IKEA and control group were randomized in order for each participant. The interaction part for the IKEA group of both tasks was structured the same (see Figure 1a).

First, participants were given the assignment, and then they were given a first, poor fitting output from the respective AI. Then, they were asked to adapt it which requires effort and were given four input options—two of them suggested improvements in relation to the problem definition, and the other two did not. Exemplary input options for the second round of the image task are provided in Figure 1b. The next output would better fulfil the problem definition but would still miss some crucial details. Thus, the participants could improve the output again. If they chose one option that did not make sense, they failed the attention check and were excluded from the experiment. Overall, they adapted the output three times until a final output was given. In addition, both tasks (text and image generation) were implemented in a version where no input selection for iteratively adapting the output was possible for the control group. Instead, the identical, final AI output was shown to the participants immediately after reading the problem definition. All images and text excerpts provided to participants throughout the experiment were created with Stable Diffusion and ChatGPT respectively, and pre-selected for this experiment by IS researchers to ensure equal experimental conditions for all participants.

The questionnaire was set up with Unipark, a survey tool emphasizing privacy and allowing us to include many different types of questions. After agreeing to a privacy policy and reading their rights according to GDPR, the experiment started with information on the task. Then, the participants were randomly assigned to one of the two groups for Task I. After each task, the participants were asked follow-up questions to measure the IKEA effect. We chose to leverage existing constructs and therefore conducted a structured literature review on the "IKEA effect" following vom Brocke et al. (2009). Consequently, we first defined the review scope. Here, we focus on research methods, especially their used constructs to measure the IKEA effect. Our goal is to summarize them. The organization is conceptual and the perspective neutral. Our audience are general and specialized scholars, and our coverage is representative. In the second step, we conceptualized the topic by reading some general literature on the IKEA effect (e.g., Marsh et al., 2022; Norton et al., 2012). Here, we also set our inclusion criteria—a quantitative study was conducted with a measurement of the IKEA effect; and exclusion criteria—no measurement of the IKEA effect. Third, we conducted the literature search. To remain as broad as possible and also include papers, especially from IS, we used the search term "IKEA effect" in WebOfScience and AIS eLibrary. This resulted in 133 and 122 papers, respectively, with one duplicate. We also conducted a backward search to find the first paper and especially the definition of the IKEA effect (resulting in 247 publications) and a forward search which resulted in one additional paper (Turel et al., 2007). Fourth, we analyzed and synthesized the literature by first excluding all papers which did not include a quantitative measurement of the IKEA effect according to their title and abstract, resulting in 33 papers. We then carefully extracted the constructs used in the papers to measure the IKEA effect.

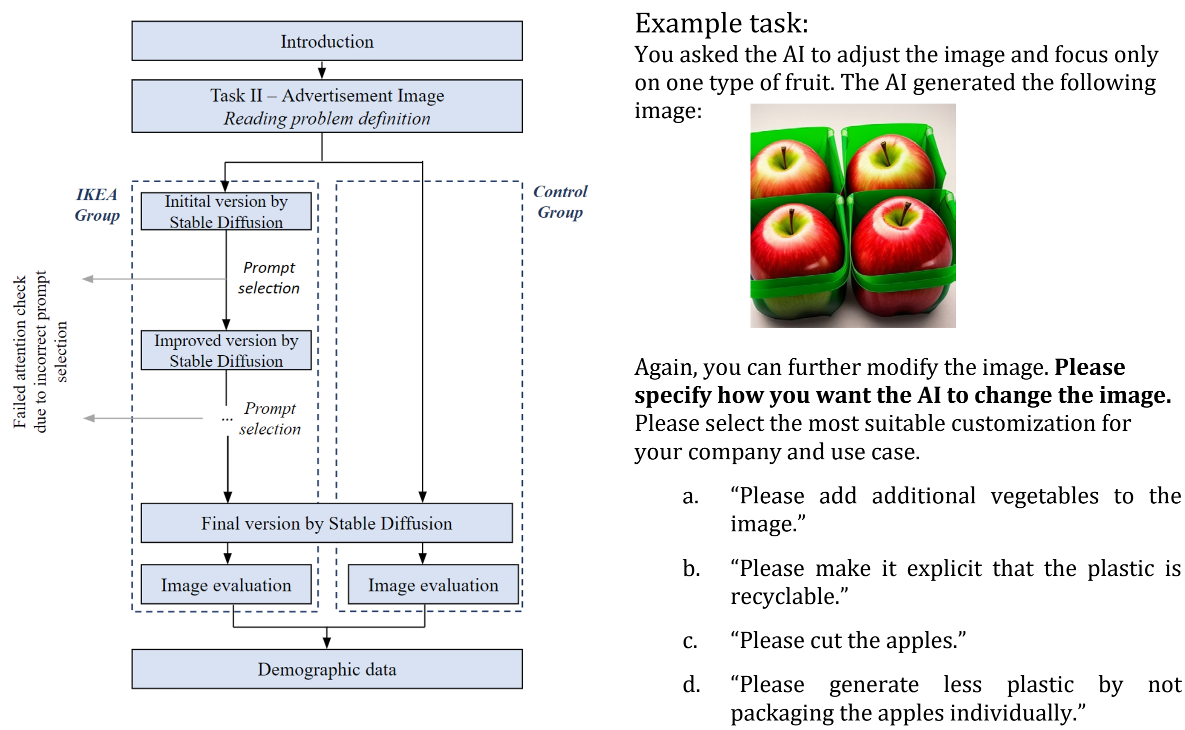

Figure 7-1 – a) Experimental Setup for the Image Task and b) Example Task With AI Collaboration; Participants Followed the Same Procedure for the Text Generation Task

This resulted in measuring the willingness to pay (one item "Are you willing to pay for the output of the AI? If yes, how much?"; T. Liu et al., 2023; Marsh et al., 2022; Norton et al., 2012), the appreciation (one item on a 7-point Likert scale: "How much do you appreciate the AI-generated output?"; Radtke et al., 2019), the perceived value (four constructs each four to five items on a 7-point Likert scale; e.g., "The quality of the AI-generated output is convincing."; Turel et al., 2007), and the behavioral intention to use the technology (two items on a 7-point Likert scale; e.g., "Assuming that I have access to this GenAI, I would use it in the future."; Turel et al., 2007). In addition, the effort is measured to prove effort manipulation between the groups was successful (one item on a 7-point Likert scale: "How much effort did you invest in creating the output with GenAI?"; Raghoebar et al., 2017). We adapted the constructs slightly to our context and translated them into German. Finally, the demographic data (age and gender) and control variables (AI experience (five items on a 7-point Likert scale, e.g., "Compared to most people, I know more about AI") following Flynn and Goldsmith (1999) and AI attitude (seven items on a 7-point Likert scale, e.g., "A majority of society will benefit from AI in the future") following Schepman and Rodway (2020)) were asked. In addition, we ensured privacy, chose high-quality constructs, and randomized the options for the participants to prevent common method bias (CMB). Also, during the experiment, validity was ensured by using existing constructs and testing the survey and reliability by using multi-item constructs if possible. In addition, we will check for both later in the analysis.

7.4.2 Data Collection and Sample

Before administering the experiment, we conducted a pretest with four experienced IS researchers. We changed some wording to make the questions easier to understand and improved some spelling mistakes. In addition, we asked three potential participants from our target group to review the experiment to get insights into unclear task descriptions and improvements. Afterwards, our target group was contacted through Prolific, a market research institute, and paid to participate in our experiment 16 €/hr. We therefore calculated the sample size following the formula that assuming the effect is small (d = 0.35), we require 72 participants to show the effect at 90% for both groups. Thus, we contacted 200 people in our target group, assuming that at least 10% would fail the attention check and choose an option to improve the text or image that does not make sense.

Overall, the experiment was expected to take 15 minutes, and the participants required, on average, 9:34 minutes. 26 participants failed the attention check and were excluded. Other data was not excluded as the attention check was comprehensive enough, every question was marked as mandatory and no straight-liners were detected. Thus, we had a final sample size of 174 participants from Germany, of whom 89 were assigned to the IKEA group and 85 to the control group for the text task and vice versa for the image task. In our sample, most participants (n = 128) were between 18 and 33 years old (74%), 32 participants were between 34 and 44 years old (18%) and the rest above 45. The gender was more balanced. We had 87 females, 84 males, and three participants identifying as divers. 68 participants are working full-time, 40 part-time, 5 are not in paid work (e.g., homemaker or retired) and 15 are unemployed but job-seeking. The other participants did not reveal their working status. Finally, participants reported an average prior AI experience (measured with five items on a 7-point Likert scale, Cronbach's alpha = .932) of 4.07 (SD = 1.45) and an average attitude towards AI (measured with seven items on a 7-point Likert scale, Cronbach's alpha = .874) of 5.19 (SD = 1.27).

7.4.3 Data Analysis and Results

We begin with checking the effort manipulation because only a perceived increase of effort can cause the IKEA effect (Mochon et al., 2012; Norton et al., 2012). For both tasks, the participants were asked about their own perceived effort (on a 7-point Likert scale) in creating the text and the image. Regarding the task of creating a company mission statement for a start-up in the healthcare sector using ChatGPT, the effort manipulation was not successful, as the IKEA group perceived the effort of the collaboration through four rounds of prompting was only slightly higher on average (M = 2.77, SD = 1.06) than the control group (M = 2.68, SD = 1.15), $t(169.02)$ [2] = -0.56, p = .290, with an effect size of d = 0.08, indicating a very small effect. The effort manipulation regarding the image creation for advertising sustainably packaging apples using Stable Diffusion on the other hand was successful, as the IKEA group reported significant higher perceived effort on average (M = 3.14, SD = 1.22) than the control group (M = 2.52, SD = 1.17), $t(170.81)$ = 3.41, p < .001, with an effect size of d = 0.52, indicating a medium effect. Thus, in the following, we will only present the results of the image creation task for which effort manipulation was successful and not the text creation task.

To assess convergent validity, we evaluated that all latent variables are above the recommended threshold of .5 for average variance extracted (AVE) and composite reliability and Cronbach's alpha exceeded the threshold of .7 (Hair Jr. et al., 2017). Table 1 further outlines that all item loadings surpass the threshold of .7 and all constructs fulfil reliability and convergent validity.

	Factor Loadings	**Composite Reliability**	**AVE**	**Cronbach's alpha**
Perceived Effort	*.768-.839*	*.785*	*.647*	*.850*
Appreciation	1.000	1.000	1.000	1.000
Quality Value	.926-.968	.971	.894	.963
Emotional Value	.829-.921	.946	.779	.938
Value-for-Money	.879-.943	.951	.830	.950
Social Value	.761-.849	.891	.671	.894
Behavioral Intention to Use	.948-.955	.950	.905	.938

Table 7-1 – Assessment of Reliability and Convergent Validity (Values Are 1.000 for One-Dimensional Constructs)

The results of our discriminant validity analysis are shown in Table 2. We verified that the square root of AVE (pictured on the diagonal; is 1.000 for one-dimension constructs) is greater than the interconstruct correlations (Gefen et al., 2000) and thus conclude that all constructs indicate sufficient discriminant validity (Fornell & Larcker, 1981).

[2] As shown later, we first performed Levene's test for homogeneity of variances. If homogeneity of variances cannot be assured, instead of a t-test a robust Welch test is performed. To ensure significant p-values the degrees of freedom are automatically adapted in the Welch test.

	Appreciation	**Quality Value**	**Emotional Value**	**Value-for-Money**	**Social Value**	**Behavioral Intention to Use**
Appreciation	**1.000**					
Quality Value	.871	**.945**				
Emotional Value	.801	.856	**.882**			
Value-for-Money	.694	.744	.762	**.911**		
Social Value	.519	.526	.574	.524	**.819**	
Behavioral Intention to Use	.713	.657	.713	.661	.600	**.952**

Table 7-2 – Discriminant Validity

First, homogeneity of variances was asserted using Levene's test. If equal variances could be assumed, unpaired t-tests were performed to test for differences between the IKEA group that invested effort into collaboration through the selection of appropriate prompts and the control group. Otherwise, Welch tests were performed. Due to the sample size of 30 or more for each group, the normal distribution of the data obtained can be assumed so that unpaired t-tests can be used as a robust measure (Stone, 2010).

H1: Willingness to pay. We found a significant difference in WTP (yes or no) for the group that collaboratively created images through prompting and the control group, $t(172) = 3.98$, $p < .001$, d = 0.48. While this effect is of small size, we further explored differences in the WTP for a monthly license. The results showed that the average price (in Euro) participants were willing to pay was significantly higher in the group that iteratively interacted with GenAI ($M = 18.58$, $SD = 62.58$) compared to the control group ($M = 5.01$, $SD = 10.68$), $t(172) = 2.02$, $p = .045$ with an effect size of d = 0.30, indicating a medium effect. Thus, we conclude that **H1** is supported by the findings of this study, and WTP for images generated by AI is significantly higher if effort is invested into collaborating with the GenAI.

H2: Appreciation. As shown by Norton et al. (2012), it is to be expected that with higher effort, the *appreciation* or *liking* of created objects increases. The findings of this study show that this bias also applies to the collaborative creation of content, such as images with a GenAI. Participants of the IKEA group reported a significantly higher average appreciation of the final image ($M = 4.88$, $SD = 1.35$) than the control group ($M = 3.71$, $SD = 1.60$), $t(172) = 5.18$, $p < .001$, with an effect size d =0.79, indicating a medium effect, even though they evaluated the same image. Therefore, we successfully demonstrated the IKEA effect in our online experiment, and the result supports **H2**.

H3: Perceived value. While appreciation or liking of an object or content would rather be described as a personal feeling, perceived value, which is a key determinant for behavioral intentions, can be measured more objectively through four constructs (quality, emotional, value-for-money, and social value) as proposed by Turel et al. (2007). We first examine H3.a through H3.d before deriving an overall assessment of whether perceived value increases with effort.

H3.a: Quality value. Participants of the IKEA group report a significantly higher quality value for the generated solution ($M = 4.77$, $SD = 1.32$) than the control group ($M = 3.70$, $SD = 1.55$), $t(172) = 4.90$, $p < .001$. In addition, this effect is d = 0.74, indicating a medium effect. We thus conclude that **H3.a** is supported, and the quality value of an AI-generated solution increases if effort is put into collaborative creation.

H3.b: Emotional value. Besides the quality value, participants of the IKEA group further reported a significantly higher emotional value, which they attached to the created final image (M = 4.94, SD = 1.19) than the control group (M = 3.76, SD = 1.40), $t(169.71) = 5.97$, $p < .001$, d = 0.91, indicating a large effect. This finding supports **H3.b** and the emotional value significantly increased with effort in our experiment.

H3.c: Value-for-money. We further asked all participants how they would rate the value-for-money if a monthly license would cost 15 euros. Value-for-money received significantly better ratings by the IKEA group (M = 4.46, SD = 1.48) than the control group (M = 3.50, SD = 1.48), $t(171.61) = 4.29$, $p < .001$, d = 0.65, indicating a medium effect which supports **H3.c**.

H3.d: Social value. Lastly, the social value of the created solution was assessed. Participants of the IKEA group attributed a significantly higher social value to their creation (M = 4.21, SD = 1.23) than the control group (M = 3.72, SD = 1.26), $t(171.95) = 2.80$, $p = .006$, d = 0.39, indicating a small effect. Therefore, in conclusion, we can say that the social value has also increased with effort, and the finding thus supports **H3.d**.

Hypothesis	**t-Value**	**p-Value**	**Cohen's d**	**Outcome**
H1: Effort—WTP	2.02	.045	0.30	Supported
H2: Effort—Appreciation	5.18	< .001	0.79	Supported
H3: Effort—Perceived Value	-	-	-	Supported by H3.a - H3.d
H3.a: Effort—Quality Value	4.90	< .001	0.74	Supported
H3.b: Effort—Emotional Value	5.97	< .001	0.91	Supported
H3.c: Effort—Value-for-Money	4.29	< .001	0.65	Supported
H3.d: Effort—Social Value	2.80	.006	0.39	Supported
H4: Effort—Behavioral Intention to Use	3.88	< .001	0.59	Supported

Table 7-3 – Hypotheses Testing

Since all four core determinants of perceived value are significantly higher for the IKEA group that invested effort in generating the image with the AI, we conclude that perceived value increases with effort, and **H3** is supported.

H4: Behavioral intention to use. Perceived value is seen as a driver for the adoption of a system or a general technology (e.g., Turel et al., 2007). We therefore examined if the behavioral intention to use the technology (GenAI for image creation) increases if participants put effort into collaborating with the technology. Interestingly, the behavioral intention to use GenAI also significantly increased for the IKEA group and thus with effort (M = 4.95, SD = 1.52) compared to the control group (M = 4.08, SD = 1.45), $t(170.49) = 3.88$, $p < .001$, d = 0.59, indicating a medium effect. We thus conclude that **H4** is supported.

7.5 Discussion

Previous research on the IKEA effect does not address whether the IKEA effect can also be observed for non-physical content and, in particular, GenAI. Especially as the results of AI are unpredictable and, to some extent, cannot be generated by humans in the same way (Berente et al., 2021; Dwivedi et al., 2023), findings from the literature are not transferable. With our experiment, we aim to determine if the IKEA effect can be found in order to be considered in the design and adoption of AI systems. Although AI seeks to automate processes, customer experience strategies of companies such as Adidas show that the contribution of effort by the customer or end user is essential so that collaboration produces valuable results, solutions, or products.

In order to determine the IKEA effect in collaborative content creation with GenAI, we divided our online experiment participants into two groups. The IKEA group was able to co-create solutions in an iterative development process that required human effort using the AI tool. In contrast, the control group received the generated AI solution immediately. Participants were randomly assigned to one of the two groups (IKEA or control group) for the image task and then assigned to the other group for the text task. The tasks were presented in a way that both groups received the identical final output, regardless of whether the task was solved for the IKEA or the control group. Unfortunately, we did not notice any difference in effort in the text generation task, which led to the IKEA effect not being evident, a finding that has also been made for physical products (Mochon et al., 2012; Norton et al., 2012). Based on feedback from the participants, we assume that the difference between the effort collaborating with AI and receiving the final output was not observed here is simply due to the fact that it is significantly more exhausting for humans to read and evaluate texts than images. Therefore, we focused our evaluation on the image generation task where effort manipulation was successful.

Various methods can be found in the literature to measure the IKEA effect. One of the most popular is the measurement of the WTP. We found a strong significant difference between the IKEA and the control group, so that H1 is supported. Thus, we can explain to a certain degree that the WTP for tools like ChatGPT is so high, because people can invest effort in the collaboration and create their own output. Likewise, we can assume H2—the appreciation of the output since a significant difference exists in the fact that the participants prefer the final version, where they themselves have invested more effort in the development process. We believe this can be due to the labor invested, causing either a signal of competence, effort justification, or feelings of ownership. Finally, we measured the perceived value of generated content through the four constructs: quality value, emotional value, value-for-money, and social value proposed by Turel et al. (2007). Also, these four constructs show a significant difference between the IKEA and the control group, so we can answer our first research question with *human effort invested in collaboration with GenAI increases the perceived value of AI-generated solutions*. Furthermore, RQ2, whether *perceived value also increases the behavioral intention to use GenAI technology* can be confirmed with H4. The behavioral intention to use GenAI tools, as in our example Stable Diffusion, can thus increase over time if the IKEA effect is exploited.

7.5.1 Contributions

With our experiment, we are able to make several theoretical contributions: First, by confirming the hypotheses, we can show that the IKEA effect can occur not only for physical products such as origami, food, or furniture (e.g., Dohle et al., 2014; Ling et al., 2020; Mochon et al., 2012; Norton et al., 2012), but also for content produced by GenAI. As (generative) AI has unique characteristics (Bauer, Hinz, van der Aalst, et al., 2021; Dwivedi et al., 2023), this effect is not self-evident. However, a sufficient difference in perceived effort must exist for the IKEA effect to manifest in using GenAI. The IKEA effect only occurs when people perceive that they have achieved and contributed to something. This can most likely also be generalized to former AI use and leads to our second theoretical contribution. With the ability of GenAI to create new content instead of solely making decisions, we show with our findings that human-AI collaboration needs to be rethought, and we add to the goals that should be pursued in human-AI collaboration. Humans now find themselves in a new role: Rather than just receiving final content or decisions, they can participate in the outcome and co-develop the result with their input. While it remains unclear what effect a prompt has on the output of GenAI, we show that collaboration is essential for human end users to value results. Here, new XAI approaches may gain importance, which face new challenges, especially with regard to image generation for example. This also changes the way humans are considered in research: With regard to the black box character, it is important to explore how humans can cooperate with AI although they may not fully understand the functionality. Even though bias is often perceived as negative, thirdly, we believe that the IKEA effect can also be leveraged. To increase the value of the outcome for users, they could be integrated artificially into the process. The IKEA effect can be utilized by developing suitable interfaces and interaction opportunities. Furthermore, our research offers practical contributions: First, we can find a higher WTP and behavioral intention to use for GenAI with human participation. This can be exploited by the providers of such services. It may also explain to some extent the high WTP for the monthly license of ChatGPT and the rapidly growing number of users. Therefore, (generative) AI providers should adapt strategies of customizing from brands like Adidas and Build-A-Bear instead of exclusively selling automation solutions to customers. These remain necessary for many tasks, especially when monotonous or dangerous tasks can be substituted. Still, also due to legal reasons, creative work or decisions in high-stakes environments are likely to remain with humans in the foreseeable future. But our findings are also valuable for the users of GenAI. Organizations planning to implement such solutions should primarily choose tasks that require collaborative approaches. Otherwise, employees might fear losing their jobs and the acceptance of these solutions could decrease significantly. By exploiting the IKEA effect, we anticipate the opposite: In collaboration between humans and machines, the perceived value of the end product increases so that the behavioral intention to use increases, which can lead to higher acceptance. Finally, we can recommend potential users of GenAI to give it a try and discover new ways of facilitating work by contributing to and shaping the outcome of AI.

7.5.2 Limitations and Future Research

While our study has added surprising insights into the realm of the IKEA effect in the context of GenAI, a few limitations must be acknowledged and further provide a basis for future research. First, the online experiment was conducted exclusively with German participants, and respective cultural norms, attitudes, and experiences may influence individual perceptions and biases. In addition, the majority of participants were between 18 and 33 years old, and the results on the IKEA effect may not be transferable to children and older people. While the results are thus relevant for most of the current users of GenAI, it is still interesting to understand how and whether the IKEA effect affects human-AI collaboration at different stages of users' lives. Second, our experimental design did not allow participants to have a "real" interaction with Stable Diffusion. While this was intentionally done to ensure uniformity in the results across participants and to trigger the IKEA effect, it is worth noting that the interaction was simulated. Nevertheless, the feedback from participants indicates that they believed they were indeed interacting with an actual GenAI, which speaks to the validity of our design. Third, the experiment focused on image generation. The realm of GenAI spans far beyond image creation, including outputs like code, music, etc. The occurrence of the IKEA effect in these contexts poses a topic for future research. Lastly, while the phenomenon of the IKEA effect was observed in the context of GenAI, the reasons for its occurrence remain unclear. Previous research on the IKEA effect has pointed to factors like the signal of competence, effort justification, and feelings of ownership as potential underlying causes. Also, simple factors such as the amount of time spent on a task could influence the perceived effort. Future research could delve deeper into understanding the specifics of why this effect is manifested, especially in the context of AI-generated content. In addition, future research should investigate whether the IKEA effect, in the context of GenAI, varies across different cultures. Moreover, our findings underline the potential influence of the IKEA effect on user perceptions. Future research should focus on how this effect can be factored into the design of GenAI tools. Incorporating psychological insights might enhance the perceived value of AI-generated solutions and human-AI collaboration in the future. In conclusion, we encourage scholars to build on this work, further unraveling the intricate relationship between humans and GenAI.

7.6 Acknowledgement

This research and development project is/was funded by the German Federal Ministry of Education and Research (BMBF) within the “Innovations for Tomorrow’s Production, Services, and Work” Program (funding number 02L19C150) and implemented by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication.

8 Paper F: A Framework for Developing Cross-Sectional Surveys

Title

A Framework for Developing Cross-Sectional Surveys

Authors

Mehler, Maren F., Ellenrieder, Sara, Turan Akdag, Merve, Wagner, Amina and Benbasat, Izak

Publication Outlet

Proceedings of the Forty-Fourth International Conference on Information Systems (ICIS 2023), Hyderabad, India

Abstract

Although the use of cross-sectional surveys is widespread in Information Systems (IS) research and related disciplines, few papers address the survey development process. In order to ensure a standardized approach, comparable and valid results, as well as to guide researchers in quantitative research methods, this paper presents a framework for the survey development process in IS. Based on a Design Science Research (DSR) methodology, the framework was derived from a structured literature review of leading IS journals and refined by three focus group discussions among IS experts. The framework includes several steps and considerations on the sample size, variable selection, their order in the survey, protection against bias, ensuring validity and reliability, and testing before administering the survey with a focus on documentation and reporting. Our framework supports quantitative research by providing a structured approach to create reliable and credible surveys.

Keywords

Cross-sectional surveys, survey development process, research methodology

8.1 Introduction

The Information Systems (IS) research domain highly values the accurate implementation and effective presentation of research methodologies. Ultimately, a sound research design and its execution are essential for any publication. Given the importance of methodology in IS research, a great body of literature exists that provides detailed guidance on how to design research projects and execute them properly. In this body of work, five overarching pillars of research methods and their respective guidance have emerged. Regarding literature reviews, Okoli and Schabram (2010), Schryen (2015), vom Brocke et al. (2009), and Webster and Watson (2002), for example, offer step-by-step instructions on how to perform literature reviews in a structured and reliable manner. For conducting qualitative research and its underlying grounded theory, Eisenhardt (1989), Myers and Newman (2007), or Urquhart et al. (2010) form the basis for almost every qualitative research project; independently of whether researchers are novices or experienced. Design science research (DSR) goes even further here, offering established papers that detail the steps a successful DSR project must go through to be relevant and rigor (e.g., Hevner et al., 2004; Kuechler & Vaishnavi, 2008; Peffers et al., 2007). Finally, Venkatesh et al. (2013), as an example, provide guidelines for mixed methods research in IS.

As a fifth research method, quantitative research, and especially cross-sectional survey studies, have been a key element of IS research for decades (Newsted et al., 1998). They allow to observe a phenomenon in a broad sample and to validate relationships statistically relatively inexpensive and easily if done right (Dinev et al., 2013; Melnyk et al., 2012; Recker, 2021). Surveys can be used for *exploration*—to become familiar with a topic; for *description*—to study behavior, opinions, processes or situations; or for *explanation*—to test theoretical and hypothetical relationships (Recker, 2021). To better define the scope of our framework, it is important to clarify our understanding of the term "survey" within the context of IS research. Compared to other quantitative methods, such as experiments, surveys are often administered to a broad sample, resulting in generalizable conclusions by using structured questionnaires (W. R. King, 2005; Mazaheri et al., 2020). In general, surveys can be categorized into cross-sectional and longitudinal surveys. The former is defined as a one-time data collection (Srivastava et al., 2015). Longitudinal surveys are characterized by surveying the same individuals at least twice, resulting in the measurement of differences over a longer period of time (Moorman, 2008). This paper refers primarily to cross-sectional surveys as they typically are part of almost every quantitative research and are regularly adopted by IS researchers. The data can originate from heterogeneous sources, nowadays especially online surveys (Che et al., 2022; Ciolkowski et al., 2003; Kranz, 2021; Mazaheri et al., 2020).

Despite its wide usage (Mazaheri et al., 2020), the literature on quantitative research (in IS) has primarily focused on two aspects: the development of constructs (e.g., Compeau et al., 2022; MacKenzie et al., 2011; Moore & Benbasat, 1991) and the statistical analysis of survey results, especially ensuring the quality of data through validity or reliability measures (e.g., Hair Jr. et al., 2017; W. R. King, 2005; Podsakoff et al., 2003; Schmitz et al., 2020), which is also the focus of the well-known website of Straub et al. (2022). However, the survey development process involves multiple steps, including constructing survey questions, designing the survey and administering the survey, before analyzing the data. To the best of our knowledge, no paper in IS has yet focused on "bridging the gap" between a researcher choosing to conduct a survey and the statistical analysis of survey data by providing detailed guidance on the steps involved in survey development, thus ensuring that the data collected is valid, reliable, and accurate. Although this process might be known by experienced researchers, providing an explicit framework can significantly improve data collection and therefore the quality of future research output. As such, the current state of the literature on quantitative research in IS remains insufficient in providing a comprehensive understanding of the research process, as the actual steps involved in the survey development process have received relatively little attention. In addition, despite surveys being used in various research fields, there also seems to be a dearth of research on the development of surveys. Since IS research is strongly methodology-driven, as many other fundamental methodology papers on, for example, structured literature review or DSR have been successfully published and IS is an interdisciplinary research area, we consider the development process of cross-sectional surveys in IS in this paper. Thus, the primary objective of this paper is to provide a framework consisting of a comprehensive set of guidelines to standardize the survey development process, thereby enhancing comparability and reproducibility across research. Especially for beginners in survey development, the framework offers a detailed and structured approach for conducting surveys, which ensures high-quality surveys that deliver accurate and reliable data. But although the concepts underlying the derived guidelines are probably known by advanced researchers, they may also benefit from this paper by using the proposed framework as a checklist that their surveys are developed to the highest standards. Finally, this paper aims to provide support to reviewers, offering a common framework for evaluating and providing better feedback to cross-sectional survey research papers. By approaching this broad range of researchers, we seek to support at all stages of the survey development process to enhance the rigor and quality of research.

We follow the DSR approach by Peffers et al. (2007) to derive a framework for the survey development process. To this end, we conduct a structured review of existing literature for the derivation of guidelines proposed in this paper to aggregate and synthesize the best practices currently available in IS research. The guidelines were first presented as part of a research project presentation to other IS researchers and then evaluated by three focus group discussions. They also expanded the framework with IS-specific tangible knowledge. Deriving the guidelines from the literature and then evaluating them through focus groups ensures that the framework is not only based on existing literature but is also practical and useful for researchers in the field of IS research. In addition, our guidelines are not intended to be prescriptive or restrictive. Instead, they provide a flexible framework that can be adapted to different research contexts and can be tailored to specific research questions that are suitable for quantitative research. For example, telephone- or face-to-face surveys can also benefit from these guidelines, but some of the guidelines may need to be adapted minimally for use in other types of survey approaches.

This paper is structured as follows: After the Introduction, which contains the problem statement and motivation of this paper as well as existing work, the Methodology is described, where the DSR approach, including the literature review and the focus groups, are described. This is followed by our framework in the results section. The paper concludes with a Discussion of contributions, limitations, and future research.

8.2 A Design Science Research Approach

To derive a structured framework for the standardized development of surveys that allows comparability and reproducibility, we chose a DSR approach similar to vom Brocke et al. (2009) or Nickerson et al. (2013). DSR allows us to iteratively derive guidelines first from literature and then improve them through a focus groups of IS researchers. Thus, we apply the DSR approach suggested by Peffers et al. (2007). First of all, following this well-established, concise approach ensures the rigorousness of the derived solution (Hevner et al., 2004). Second, while we aim to deepen our understanding of survey development in the IS research community, it is a key objective to ensure comprehensibility as well as broad practicability of the derived guidelines underlying our framework. As many surveys investigate real-world problems, the applied DSR methodology is suitable for developing an artifact in the form of a framework that allows for broad applicability (Baskerville et al., 2015; Venable & Baskerville, 2012). The derived framework defines clear guidelines for each step in the development process of a survey and thus provides a Level 2 contribution in terms of the DSR contribution types by Gregor and Hevner (2013). Example artifacts for Level 1 contributions are specific solution instantiations, while Level 2 contributions are, e.g., constructs, methods, models, design principles, and technological rules, and Level 3 contributions include the derivation of new design theories (Gregor & Hevner, 2013).

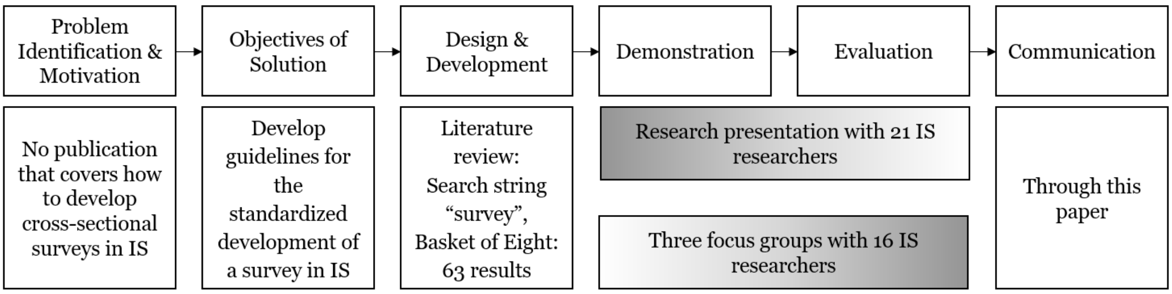

Figure 8-1 – DSR Methodology for Our Study in Accordance With Peffers et al. (2007)

Figure 8-1 visualizes our approach, which consists of six distinct phases, namely 1) problem identification & motivation, 2) objectives of solution, 3) design & development, 4) demonstration, 5) evaluation, and 6) communication with possible starting points at 1) to 4) (Peffers et al., 2007). We began the design cycle in phase 3) by first reviewing existing literature on surveys in the Basket of Eight journals (*European Journal of Information Systems*, *Information Systems Journal*, *Information Systems Research*, *Journal of the Association for Information Systems*, *Journal of Information Technology*, *Journal of Management Information Systems*, *Journal of Strategic Information Systems*, and *MIS Quarterly*) to account for rigor in the design process (Hevner et al., 2004). A detailed description of the performed journal review is provided in the next section. Besides identifying the lack of clear guidance for developing surveys for phase 1, we were able to define the key objectives for the framework in process phase 2. Based on the surveys that were published in the Basket of Eight journals, guidelines for all process phases of the development phase of surveys were derived. Standardization and broad applicability of guidelines were the key objectives during this phase 3. Peffers et al. (2007) suggest demonstrating the artifact in phase 4 before evaluating it in phase 5. We demonstrate and subsequently evaluate the guidelines in different settings to ensure comprehensibility and effectiveness while also obtaining in-depth qualitative feedback from IS researchers. In a presentation of our research project with 21 IS researchers the identified research gap as well as derived guidelines were shown before an open discussion was conducted on the relevance, comprehensibility, and completeness of the framework. The main focus of the research presentation was to demonstrate the guidelines to a large group of researchers, who were then asked about previous their experiences with surveys and then were invited to participate in several focus groups according to their level of expertise. Focus groups allowed us to evaluate how selected groups of researchers with different prior experiences evaluate the derived guidelines (O'Nyumba et al., 2018). The DSR approach by Peffers et al. (2007) requires communication (phase 6) of the results obtained to appropriate audiences. As we primarily focus on surveys conducted in IS research, this paper allows us to present the conducted design cycle and final framework to IS researchers who can build their future research on the blueprint provided in this paper.

8.2.1 Literature Review

To derive the initial set of guidelines, we conducted a literature review following the five steps according to the approach suggested by vom Brocke et al. (2009). First, we **defined our review scope** using Cooper's (1988) Taxonomy which consists of six characteristics guiding researchers to specify their scope. Our *focus* is on research methods, as it is the case with quantitative surveys, and our *goal* is to integrate and combine the literature into a framework. We *organize* our review conceptually and take a neutral *perspective*. Our *audience* are both specialized and general scholars in IS and we provide representative *coverage* by focusing on articles in the Basket of Eight journals that ensures high quality as they have already passed an extensive review process. Second, we **conceptualized the topic**. We agreed on the most general search string "survey" to remain as broad as possible. Since we focus on cross-sectional surveys, we chose the following inclusion criteria: 1) At least one cross-sectional survey was conducted in the paper, and 2) the paper provides sufficient insights into the survey itself or its development process. We did not consider papers that focus on research methodologies for the literature review but read them to substantiate our framework. Third, in our **literature search**, we used the search string "survey" in WebOfScience which allowed us to specifically selecting only articles from the Basket of Eight journals as they are extensively reviewed and accepted as high-quality journals (Polites et al., 2012). This resulted in 741 articles. We read and applied the inclusion and exclusion criteria on the papers according to the publication year by first screening the titles and abstracts before reading them (Schryen, 2015). To reach saturation we **analyzed and synthesized** papers according to their publication year adding more papers until no paper provide any new information on the survey development process. Thus, we started with the 23 papers from 2022 by scanning the titles and abstracts. We excluded three papers that were not conducting a survey. The remaining 20 papers were then read carefully. Based on them, we developed a list of coding categories, such as used constructs, included control variables, measurements against biases, and types of testing the survey, which we then applied to the selected papers. Afterward, we used the categories to derive a first draft of guidelines. We already noticed saturation in the guidelines after analyzing the papers from 2022. We then extended our guidelines with articles from 2021, applying the same procedure as before, again excluding seven papers as they did not conduct a survey, resulting in 27 articles. As hardly any new aspects for the guidelines emerged, we chose as representatives two random articles from each of the last 10 years to finalize the guidelines. Finally, we used a backward search to include relevant methodology papers—which are not counted here as they are only used in the guidelines and not to derive them—and we did not perform a forward search as we only needed a representative set of survey papers. Overall, we included 65 (20 from 2022, 27 from 2021, 2 each from 2012-2020) papers in the literature review. A **research agenda** as the last step will be included in the discussion as an outlook on future research ideas.

8.2.2 Focus Groups

To evaluate the framework, ensure the relevance and correctness of the guidelines, and for additional input, we conducted three focus groups. A qualitative methodology is beneficial, as the knowledge collected is not shared in public in the form of papers but is anchored in the interviewees based on their experience. The main advantage of focus groups compared to interviews is that a discussion takes place between the participants and different opinions influence the result (Gibson & Arnott, 2007). In addition, Gibson and Arnott (2007) recommend focus groups for evaluating the utility, quality, and efficacy in DSR. The three focus groups took place in March 2023 and consisted of five to six IS researchers who knew each other at least to some extent and with different research interest such as privacy, AI or healthcare. Six experts (one postdoc and five researchers at the end of their Ph.D. studies) with at least four (mean 10) already conducted surveys were invited to the first focus group, since we expected the highest amount of input here. The participants were asked to speak openly about their experiences and advice with surveys along the survey development process (see Figure 8-2) to improve and expand our guidelines. Following this, we asked five proficient researchers (in the middle stages of their Ph.D. studies) who had conducted at least one, but maximum two surveys to give input to each guideline by its title, hoping to gather in particular their difficulties and learnings to gain feedback on the usefulness, completeness and clarity of our guidelines. Finally, we evaluated the guidelines for ease of use, understandability and practicality with five inexperienced (no surveys conducted) researchers at the beginning of their Ph.D. studies by summarizing the guidelines. Between each focus group, the guidelines were adapted to the new insights. Following Tremblay et al. (2010), we conducted an exploratory focus group with the six experts ("Ex") and a confirmatory focus group with the five inexperienced participants ("I"). The focus group in between them consisted of five proficient participants ("P") was both exploratory and confirmatory. In order to maximize the feedback from the participants and to minimize the length of the interviews to avoid fatigue, we chose relatively small groups compared to Tremblay et al. (2010)'s suggestion of 6-12 participants. The same author moderated the focus groups and afterwards transcribed them for analysis. The discussions were, on average, 59 minutes. Finally, the focus groups were coded during content analysis using MAXQDA software (VERBI Software, 2022) by the authors to minimize the influence of individual researchers and to ensure objectivity. For this purpose, we applied the guidelines derived from the literature to the transcripts (deductive) and simultaneously identified new, context-specific aspects for the framework through an inductive approach (Mayring, 2000). From the focus groups we were able to refine the guidelines regarding tacit knowledge, for example how to order the questions in a survey or that it is important and normal to iterate back anytime during the development of a survey.

8.3 A Framework for Cross-Sectional Survey Development

In the following, the guidelines of the survey development process derived from the literature and enriched and validated by the focus groups are presented. In order to clearly distinguish the results of the literature search and the focus groups, the sources are always provided. As our focus is on the methodological part of a paper, the guidelines start after the research question has been defined and a theory—used synonymous with model or framework and describes the presumed relationships between the constructs—has been selected. The theory can be self-developed, e.g., derived from interviews, or it can be an established theory that has been extended by self-selected variables. Thus, the finding of a theory and the development of hypotheses are not included in this paper. However, especially guidelines 1, 2a, 3 and 4 (in Figure 8-2) are dependent on the chosen theory. In addition, as many papers exist on how to analyze data and this is depending on the specific research questions, our guidelines end before the analysis.

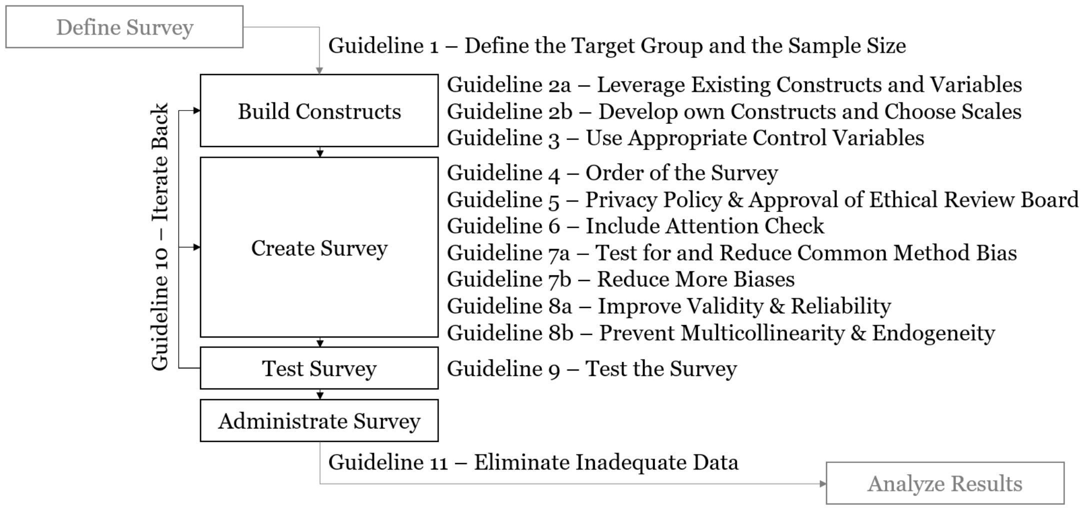

Figure 8-2 – Framework for Cross-Sectional Survey Development

Figure 8-2 gives an overview of the framework, including the guidelines and their order for the development of a survey.

Guideline 1—Define the Target Group and the Sample Size

First, the target group must be determined, because elements of other guidelines such as the type of attention check (see Guideline 6) or the phrasing of the items (see Guideline 2) are based on it (Expert 5—Ex5). The target group depends on the research question and chosen theory and, in particular, three alternatives are used: A representative group (e.g., according to age and gender in a country (J. Wang et al., 2016)), a specific group (such as employees of companies (Kranz, 2021), managers of different companies (Liang et al., 2022), or members of a platform (Abhari et al., 2022)) or a student sample as a convenient sample (Serenko & Turel, 2021). In the IS literature, all three variants are used, but mainly specific groups.

The sample size ranges from 100 (J. Zhou et al., 2022) to 1281 (Rai et al., 2022) across the papers. Rules of thumb and formulas exist to calculate the minimum numbers; however, no paper includes them or any rationale on the number of participants. As the sample size is also depending on the requirements during the analysis, we will only give the suggestion to follow Maier et al.'s (2023) recommendation of the website from Wang and Ji (2020, http://riskcalc.org:3838/samplesize/) to calculate the sample size for cross-sectional surveys.

Furthermore, the researchers must consider the form of recruitment. Market research institutes are frequently approached for this purpose, as they offer representative samples and Steelman and Hammer (2014) did not find a statistical disadvantage in acquiring participants through platforms such as online crowdsourcing markets. However, these can quickly become very expensive (Steelman & Hammer, 2014). If a specific target group is being surveyed, researchers often have to recruit participants via contacts, e-mails, LinkedIn, etc. A student sample, on the other hand, is easy to recruit, as students can be asked to participate in courses, but has the most issues including low heterogeneity thus low representativeness (Ex1). In this context, also the question of compensation arises. At a market research institute, participants are often paid a fixed amount depending on the duration of the survey, a few dollars/euros. Students or members of platforms often receive an incentive in the form of a fixed amount or the prospect of a possible prize in gift cards or even products such as iPads to increase response rate (Feng et al., 2022; Melnyk et al., 2012). Employees or managers of companies often do not receive an incentive.

Therefore, we propose: Document **(Guideline 1.1—G1.1)** the target group and why it was chosen. State **(G1.2)** how many potential participants exist, the sample size calculated and according to which formula or rule of thumb, how many participants have been contacted, and how many have participated in order to calculate the response rate. Specify **(G1.3)** which type of recruitment was chosen and **(G1.4)** whether incentives were offered.

Guideline 2a—Leverage Existing Constructs and Variables

From the chosen or developed theory measures for the survey must be derived. The existing, often abstract constructs from the theory are operationalized through specific variables (Lund Research Ltd, 2012). Overall two approaches exist: First, leveraging existing variables, including their items—variables are usually measured through multiple similar questions, which have the same scales, and second, developing new constructs and variables (see Guideline 2b). Relying on established constructs and variables has the advantage that their validity has already been proven and results can be compared across several papers (Ex1; Ex3; P1; P2; P3; P4; P5). If multiple variables exist for the selected construct, we recommend choosing the best known, most suitable and most recent one. Here, the relevance should be assessed (Compeau et al., 2022). Picking individual items from various existing variables is not recommended. Often, variables need to be adapted to the survey context. If only minor changes are needed, the items can be modified. For example, Cichy et al. (2021) chose the privacy concern collection measurement from Smith et al. (1996). Originally the item is "It usually bothers me when companies ask me for personal info." which is adapted to "It usually bothers me when the company asks me for personal driving data." The adaptation should not result in a completely new item (P4). To determine whether more substantial changes are required, Compeau et al. (2022) describe a four-step approach.

Existing (occasionally also self-created) items have to be translated sometimes. Based on Brislin (1970), the following procedure has been used by many researchers (e.g., Maier et al., 2021): The questions are translated into the target language by a bilingual individual and then translated back by a second bilingual individual. If both versions in the original language are now considered to be without differences in meaning, a pretest should be carried out in the target language. Afterward, the questions should be re-evaluated by bilingual persons, one group seeing only the original, one group seeing the translation, and one group seeing both. In case no differences in the results are detected, the translation can be adopted (Brislin, 1970). This process is often carried out in a condensed version, for example by relying on professional translators to translate the text into the target language and then re-translating the text back into the original language (e.g., Hovav et al., 2021; L. Wang et al., 2022; E1).

Therefore, we propose: Document **(G2a.1)** the source of constructs and variables, **(G2a.2)** their relevance, e.g., following Compeau et al. (2022), **(G2a.3)** the extent to which they have been adapted to the new context and **(G2a.4)** how there were translated. Best practice is to report all constructs, variables and their items in the paper.

Guideline 2b—Develop Own Constructs and Choose Scales

Adequate constructs and variables may not always exist in the literature. However, developing constructs should be the exception as it is a time-consuming process (Ex2; Ex3; Ex5). The development of new constructs usually follows MacKenzie et al.'s (2011) ten steps scale development process from the definition of the construct over the generation of items, the collection of data for testing, the elimination of items, and the validation of the development of norms (e.g., Hoehle & Venkatesh, 2015; Riemenschneider & Armstrong, 2021). They also elaborate on the difference between formative and reflective measurement models and thus provide a solid guideline for the development of constructs.

Variables can consist of multiple (preferred) or a single item, but can be also answered through open-ended, ranking or many more types of questions. Churchill (1979) recommended the use of multiple items because this results in higher reliability (see Guideline 8a) and is suitable for structural equation modeling. But it is also possible to use single items, often done by practitioners because of their simplicity. They are accepted in marketing research if they are doubly concrete constructs (Bergkvist & Rossiter, 2007). To measure items, in IS and related areas, mostly Likert scales have been adopted. Here, an item is given as a statement and the respondents have to give their attitude on a range from "strongly disagree" over "neutral" to "strongly agree" (Flamer, 1983; Likert, 1932). Commonly the range has seven steps (e.g., Cichy et al., 2021; Lin et al., 2021), but also five (e.g., Ke et al., 2021) and ten (e.g., Rai et al., 2022) are used. Regarding the choices of the best length and whether the scales should be even or odd, no clear direction is given (Leung, 2011). For the length, the literature argues in favor of long scales because of the larger variety of options for the participants so they are supposed to find their true value more likely (Joshi et al., 2015). On the other hand, an overly fine scale can be overwhelming (Matell & Jacoby, 1971). Sometimes it can be beneficial to change the wording and use semantic differential scales (Chin et al., 2008). This results in more variable and flexible scales and shorter items, e.g., instead of "Using the system enhances my effectiveness" on a scale from strongly disagree to strongly agree the item would be "This system is …" ranging from efficient to inefficient (Chin et al., 1996, p. 690). In addition, there are other scales that can be used, such as the Stapel scale or the Thurstone differential scale (Albaum, 1997). If an item comes with a scale, we recommend using it. Other types of questions are depending on the variable. Open-ended questions can, for example, be used to find reasons in participants behavior and analyzed qualitative or ranking, for example, in conjoint analyses.

Therefore, we propose: Document **(G2b.1)** which constructs are newly developed and **(G2b.2)** the process of their development. In addition, we recommend describing this process in a new paper or at least as detailed as possible including testing of the new constructs in the current paper. Additionally, **(G2b.3)** list which scale was used for all variables.

Guideline 3—Use Appropriate Control Variables

Control variables or if they refer to a person demographic data are included to account for effects that may confound the relationship between independent and dependent variables and thus to rule out alternative explanations (Atinc et al., 2012). This also allows identifying potential group differences (Kung et al., 2015). Furthermore, demographic data can be used to determine whether certain quotas are already met, e.g., age distribution in the country under consideration (E1; P1).

If the survey focuses on individuals, in particular, age and gender are included in the survey, as well as income, employment status, work experience, and sometimes marital status and the number of children. Depending on the topic, more specific variables can be utilized (e.g., Che et al. (2022) measured the experience in online shopping). Demographic data is often presented in the methodology section. When collecting demographic data, privacy must be ensured to avoid individuals being identifiable (see Guideline 5). For surveys focusing on companies, control variables such as size and industry are measured in particular. In addition, as such surveys require one person to respond on behalf of the company, demographic data for this person is often included as well *"because the demographic data are always collected"* (Ex3; P1; P3; P4). However, measuring every control variable that might have an influence is not the solution.

Carlson and Wu (2012) even recommend omitting control variables if there is a lack of evidence to support their inclusion. Bernerth and Aguinis (2016) presents a decision tree: Only if the variable is likely to have an influence that has already been examined and confirmed by other researchers and offers an alternative explanation, the inclusion of the control variable is reasonable. However, the focus group participants stated that they prefer to measure more control variables than to realize later that they missed one (P1; P4).

Therefore, we propose: Document **(G3.1)** all control variables and demographic data that are collected during the survey. Include if possible **(G3.2)** a valid rationale why the control variable is measured and **(G3.3)** show an overview of the values and their influence on the model in the analysis.

Guideline 4—Order of the Survey

After defining the measurements, they must be arranged within the questionnaire "*for the quality (…) of the survey*" (Ex6; Drury & Farhoomand, 1997). Regarding the order of the survey, no explicit statements can be found in the examined literature. A survey should start with a privacy policy focusing on confidentiality and anonymity and information about the survey (see Guideline 5). In addition, important definitions should be included at the beginning to achieve a common understanding by using not only plain text, but also images (Kranz, 2021; P4). The variables can be structured based on the relationship between the constructs from in the theory, placing mediators at the end (Ex6; P2; P5). However, sometimes it is useful to measure the dependent variable first to avoid influence by other questions (e.g., questions about actual behavior can be influenced if people are asked how they handle privacy) (Ex1; P3). Regarding control variables, they are often included as first (e.g., Wiener et al., 2021), which makes it possible to check directly if quotas (e.g., balanced sample size according to gender) have to be met (P3), or last part of the survey (e.g., Abhari et al., 2022), "*because it's something that tends to bore the participants*" (P3). Finally, it is useful to offer an open feedback box in the end (Ex2).

In online surveys, the number of questions on a page should be limited so that participants only have to scroll once (P1; P4). Each page should consist of approximately the same number of questions or take the same amount of time (P2; P3). In addition, a page should be thematically independent (Ex5; P1). A progress indicator helps the participants to maintain an overview and remain motivated (P3; P4). Furthermore, a back button should not be included if the participant should not change their original opinion (Ex5; P1; P2). Moreover, making questions mandatory has the advantage that users must not be eliminated at the end because of missing data (see Guideline 11), but should be avoided with private questions (P1; P2). Finally, it is useful to block questions of one variable with Likert scales in a matrix (P1), but reversed questions in a matrix should be placed carefully to avoid that participants overlooked them (Ex2; E4). Finally, the questions should also not be overly complex and long (P3).

At the variables' and items' levels, a further question arises whether they should be asked in the same order or randomly for each participant as it can influence the reliability and construct validity (see Guideline 8a). To reduce common method bias, a recommendation is to randomize the order of the survey questions (Podsakoff et al., 2003; P. N. Sharma et al., 2022; L. Wang et al., 2022). Wilson et al. (2021) recommend not to use intermixed approaches (mixing items from different variables), but to order the variable blocks randomly.

In this context, the length of the survey needs to be briefly addressed. Although no specific rules exist, the length of the survey has an impact on the response rate and can therefore threaten the validity of the survey (Chin et al., 2008; Melnyk et al., 2012). Additionally, the reliability of later questions may decline as participants become increasingly fatigued. The experts from our focus groups recommend a survey length of 10 to 20 minutes.

Therefore, we propose: Document **(G4.1)** in which order the survey components were arranged. If an online survey was conducted, **(G4.2)** describe the settings in the tool used to set up the survey. If the order of questions is random, **(G4.3)** describe how it was set for different participants, and indicate **(G4.4)** how long the survey takes on average.

Guideline 5—Privacy Policy & Approval of Ethical Review Board

To ensure privacy, it is important to include information about the storage and use of the collected (demographic) data for the participants and inform them about their rights (e.g., GDPR and that the survey is voluntary so they can leave anytime) (P1; P2). Often, universities provide information on how to implement privacy policies on their website (P3). Especially in psychology, ethical principles are ensured by the approval of an ethical review board. This should be considered depending on the topic and questions, e.g., when collecting sensitive such as medical data from the participants.

Therefore, we propose: Document **(G5.1)** if a privacy policy is provided and **(G5.2)** if an ethical board approves the survey.

Guideline 6—Include Attention Checks

Since participants may become inattentive during surveys, we recommend including at least one attention check to improve data quality. A specification on which attention check was included was rarely provided in the literature. Only the following example was found: Participants are asked to “please select strongly disagree” (Wiener et al., 2021).

Abbey and Meloy (2017) provide an overview of different types of attention checks along with the simplicity of implementation and objectivity of these checks: Logical statements, directed queries (including the example above), manipulation checks, open-ended queries, infrequency, response pattern/ time, honesty check, reverse scaling, memory recalls, and outlier detection. Except for response pattern/ time and outlier detection, attention checks must already be planned during survey creation. An attention check should be placed in the middle of the survey (P4). It should be noted that if you ask a “Please check strongly disagree” question, it is possible that the questions surrounding it will be influenced by the strong statement (Ex3). The advantage of attention checks is that they can be used to easily eliminate data sets where participants have made little effort (see Guideline 11), thus ensuring high data quality and that the survey reflects the true behaviors, attitudes, and beliefs of the participants (Abbey & Meloy, 2017). Especially in the context of online surveys using tools like Amazon Mechanical Turk, the inclusion of attention checks is recommended (Paas & Morren, 2018). However, Hauser and Schwarz (2016) discovered that Amazon Mechanical Turk respondents learned how to answer these attention checks.

Therefore, we propose: Document **(G6.1)** which type of attention check was used and **(G6.2)** how many participants failed it and were thus eliminated.

Guideline 7a—Test for and Reduce Common Method Bias

Often, surveys are checked for the common method bias (CMB or common method variance) (Fuller et al., 2016; Podsakoff et al., 2003). CMB can occur when data are collected from the same source. A faulty measurement caused by a bias can lead to incorrect results and thus flawed conclusions. Different causes for the CMB exist: “a common rater, a common measurement context, a common item context, or from the characteristics of the items themselves” (Podsakoff et al., 2003, p. 885). An overview of options for control in design or in statistics, i.e., after the survey has been conducted is provided by Podsakoff et al. (2003). The focus here will be on the options in the design: Whenever possible, the independent and dependent variables should be collected from different sources in order to minimize the common influence on them, but collecting data from the same respondents does not automatically result in CMB. When conducting longitudinal surveys, it is possible to query the two types of variables in different surveys (Gong et al., 2021; L. Wang et al., 2022). In case it is not possible to collect the data from different sources, a separation—temporal by ordering the survey questions (see Guideline 4), psychological by cover story, methodological or proximal by different response formats—can be arranged. Another possibility to reduce CMB is to guarantee anonymity to the participants (see Guideline 5) and to assure that there are no right or wrong answers. Especially to reduce priming effects it is recommended to vary the order of questions. Using a marker variable (e.g., "Coffee is important in my life") in the research model could also support determining CMB. If no high correlation with other variables can be found, the possibility for CMB is low (Maier et al., 2021; Ex2; Ex3; P3). Finally, by ensuring high quality constructs and testing them, CMB can be reduced (P3) (see Guideline 2). Cram et al. (2022, p. 440f.) summarize "keeping questions simple, focused, and concise; avoiding double-barreled questions and conceptual dependence between dependent and independent variables; using randomized items and reiterating respondent anonymity along with the exclusive research purpose of our study". This is extended by Jordan and Troth (2020) with higher item clarity and including reverse items or balancing positive and negative items.

After the data collection, the data can be additionally tested for CMB. As these tests are usually part of the analysis rather than the methodology, only two will be mentioned and briefly described here (Maier et al., 2021): First, Harman's single factor test, which indicates whether the majority of the variance can be explained by one single factor. Second, extremely high correlations ($r > 0.90$) are an indicator of CMB.

Therefore, we propose: Document **(G7a)** how the survey was protected against common method bias.

Guideline 7b—Reduce Additional Biases

Surveys are vulnerable to other types of biases as well. Therefore, again, the survey should be protected against certain biases in the survey design:

Social desirability bias: When the survey is distributed, social desirability bias can be reduced by assuring anonymity (see Guideline 6) or by sharing the survey on a survey website (Gong et al., 2021).

Non-response bias (or response bias): When conducting surveys, usually not all of the requested participants respond, which results in a subsample of the total sample. Those who do not participate may differ from the responding participants in important aspects. Especially the attractiveness of the study can increase the number of respondents (Ex4). Also, Armstrong and Overton (1977, p. 396f.) suggest the following options: Comparison with known values for the population, subjective estimates, or extrapolation methods. However, the literature shows mostly a-posteriori methods as test for significant differences using t-tests between early respondents and late respondents or ANOVA on the measurements and control variables (e.g., Kranz, 2021). Furthermore, sometimes it is the intention that only certain groups participate: *"It's completely reasonable that some groups won't be included. If you're measuring AI acceptance, it doesn't make much sense to ask people who don't use the Internet, because they'll never be in touch with AI"* (Ex2).

Other biases, such as self-selection bias, exist, however, we could not find protections against them.

Therefore, we propose: Document **(G7b)** how the survey was protected against the respective biases.

Guideline 8a—Improve Validity and Reliability

Similar to the common method bias, validity and reliability are often checked after the data has been collected, but some approaches exist to ensure them while implementing the survey. For more information on validity and reliability, we recommend Straub et al. (2022).

Validity is achieved if a measure is equal to the true value (Churchill, 1979). Different types of validity can be found in the literature: Conclusion validity indicates the extent to which a relationship between two constructs is random (Schmitz et al., 2020). Construct validity indicates to what extent the construct measures what it is supposed to measure (Peter, 1981; Voorhees et al., 2016). Content validity indicates the extent to which the variable captures what it is supposed to measure (Schmitz et al., 2020). Convergent validity indicates the extent to which the items of a variable positively correlate and discriminant validity if a variable is empirically distinct from different variable (Hair Jr. et al., 2017). External validity indicates to what extent the results are generalizable "across different measures, persons, settings, or times" (W. R. King, 2005, p. 882). In general, surveys provide high external validity (Ciolkowski et al., 2003) and are therefore easily generalizable. Internal validity indicates to what extent causality is present (Schmitz et al., 2020).

When creating the survey, the following is important to improve validity: Internal validity can be ensured by using well-established measurements and collecting data carefully (Maier et al., 2021; P2; P3). Content validity requires careful development of the constructs, especially by using card or Q sort during the scale development (Moore & Benbasat, 1991). In addition, it can be enhanced by pre-testing the instruments and receiving professional advice from experts (Hovav et al., 2021; Schmitz et al., 2020). Construct validity can also be increased by testing the survey (see Guideline 9). Many tests and procedures exist that can be performed a posteriori to check if the measures are valid. Such as the average variance extracted (AVE) for convergent validity of reflective constructs with a threshold of 0.50 (see Fornell & Larcker, 1981; Hair Jr. et al., 2006, 2017; and e.g., Srivastava et al., 2015; J. Zhou et al., 2022). For discriminant validity first the cross-loadings can be compared, where higher loadings for another variable would violate validity (e.g., Abhari et al., 2022); second the Fornell-Larcker criterion where the square root of AVE should be higher than the variables' correlations (e.g., Srivastava et al., 2015) or third—as both measures have weaknesses—the heterotrait-monotrait ratio (HTMT) where the value should be below the threshold of 0.90 to suggest discriminant validity (Hair Jr. et al., 2017; Henseler et al., 2015; e.g., Maier et al., 2021).

Reliability "describes the extent to which a measurement variable or set of variables is consistent in what it is intended to measure across multiple applications of measurements" (Straub et al., 2022). To ensure reliability prior to the data collection, only one action could be identified: Use multi item variables (Bergkvist & Rossiter, 2007; Churchill, 1979). To test reliability of the variables after data collection in general two measures are used: Cronbach's alpha with a threshold of 0.70 (see (Hair Jr. et al., 2006)—(e.g., Cichy et al., 2021; Dinev et al., 2013; Hoehle & Venkatesh, 2015) or composite reliability (CR) also with a threshold of also 0.70 (see (Hair Jr. et al., 2006) and (e.g., Abhari et al., 2022; L. Wang et al., 2022)). Both indicate internal reliability if the value exceeds the threshold, but Cronbach's alpha is more likely to underestimate and CR to overestimate the internal reliability (Hair Jr. et al., 2017). Therefore, it is best to calculate and report both.

Therefore, we propose: Document **(G8a.1)** how validity and **(G8a.2)** reliability were ensured, improved and tested. Be careful to differentiate between reflective and formative constructs.

Guideline 8b—Prevent Multicollinearity and Endogeneity

Multicollinearity indicates a correlation between several variables. Multicollinearity is problematic for formative constructs, but desired for reflective constructs (Petter et al., 2007). When selecting variables (see Guideline 2), it is important to consider whether there is conceptual redundancy. Multicollinearity can be determined afterwards, for example, by calculating the variance inflation factor (Lin et al., 2021; Maier et al., 2021). Here the rule of thumb is that if the factor is > 5 moderate or > 10 high multicollinearity is present (see Craney & Surles (2002)).

Endogeneity results if an independent variable and the regression's disturbance term are correlated which can bias the collected data and reduce validity (Sande & Ghosh, 2018). Endogeneity arises from measurement errors, simultaneity, or omitted variables (Sande & Ghosh, 2018). The following four options are recommended to address endogeneity: First, control variables (see Guideline 3) can reduce bias due to omitted variables. Second, when possible, binary variables can also be adopted as independent variables instead of Likert scales. Third, collect instrumental variables from other data sources if common method bias is the cause of endogeneity (see Guideline 7a). Fourth, measure the instrumental variables at an earlier time point than the endogenous independent variable if simultaneity is the issue (see Guideline 4) (Sande & Ghosh, 2018). From the literature review a Durbin-Watson test (e.g., Rai et al., 2022) is one approach to determine endogeneity after data collection. Others can be found in Sande and Ghosh (2018).

Therefore, we propose: Document **(G8b.1)** how multicollinearity and **(G8b.2)** endogeneity were prevented.

Guideline 9—Test the Survey

Every survey must be tested to identify flaws before administering, as small errors can quickly occur, such as a question being phrased unclear in meaning, because it was translated incorrectly. In addition, the flow, order of the questions (see Guideline 4), and timing are also factors that can be improved with testing a survey (Bolton, 1993). When preparing a pretest, five steps are stated by Hunt et al. (1982): First, determine which parts of the survey should be tested. We recommend testing the whole survey. Second, the method for the pretest must be determined. Depending on the type of pretest, it can be performed qualitatively or quantitatively (Kranz, 2021). Third, it must be specified who will conduct the pretest. Fourth, decide who will participate in the pretest. The following procedures were found in the reviewed literature: (1) Pretest with colleagues, who are not part of the research team, but experts in the research area (e.g., Cram et al., 2022; Wiener et al., 2021). Here, especially the wording and survey flow (Wiener et al., 2021), readability, understandability, and realism of the instrument (Trinkle et al., 2021) can be adjusted. (2) Pretest with students as a convenience sample to check the validity and reliability of the scales (e.g., Serenko & Turel, 2021; P. N. Sharma et al., 2022; Trinkle et al., 2021; L. Wang et al., 2022). (3) Pretest with potential participants (e.g., Hovav et al., 2021), for example, with a purchased sample (Seymour et al., 2021). The focus group participants preferred a combination of academic experts followed by potential participants (Ex1; P1; P2; P4). And fifth, how large the sample should be. Depending on the previous four steps and the sample size, a pretest sample between 5—for feedback from researchers (e.g., Cram et al., 2022) or qualitative feedback (e.g., Kranz, 2021)—

and 50 participants for a more convenient sample (e.g., L. Wang et al., 2022) or for the target group of the main survey (e.g., Abhari et al., 2022), should be sufficient (Hunt et al., 1982).

Quickly forgotten during testing is to ensure that basic aspects such as all branches are functioning (P1), the survey works on every common device, multiple languages are equivalent (P2), but also that the variables are properly implemented (Ex1; Ex2; Ex6). This can be achieved for example via a comment function of the tool (Ex1; P2) or open feedback in the survey (Ex6). Furthermore, self-developed constructs should always be tested extensively (see Guideline 2b).

Therefore, we propose: Document **(G9.1)** which part of the survey was tested and **(G9.2)** how. Specifically address **(G9.3)** which test group was contacted, **(G9.4)** how many participants tested the survey, **(G9.5)** in which order and **(G9.6)** with which focus. In addition, **(G9.7)** provide information on the changes after each test.

Guideline 10—Iterate Back

In particular after testing the survey, but also any time before administering the survey, it is possible to iterate back and adjust previous decisions, e.g., to adapt items if they are difficult to understand (P4). Iterating back into the survey development is not a one-time step, but should be repeated several times. P2 believes this guideline is the most important: "*You should take as much time as possible to optimize the survey before you start to collect the data, because it hurts less than if you haven't done it and then collected data that you can't use because you have to throw everything away and then start all over again.*" (P2).

Therefore, we propose: Document **(G10)** what adjustments were made when iterating back in the survey development process.

Guideline 11—Eliminate Inadequate Data

Some data sets need to be removed to avoid altering the results of the survey, since surveys are prone to human mistakes. The following data sets can be removed or are removed automatically by survey tools (e.g., Lin et al., 2021): Data sets that are incomplete (e.g., if single items have not been answered (P2)), where the attention check was failed (see Guideline 6) (P4), that do not meet a requirement (e.g., Riemenschneider and Armstrong (2021) eliminated 29 data sets of individuals who are not in the IS field, as this is the target participant group; however, it is recommended to sort these persons out with a question at the beginning), containing inconsistent information (e.g., years in IT > age (Riemenschneider & Armstrong, 2021)), where the response time is too short or too long (e.g. Wiener et al. (2021) sorted out if the minimum time (based on preliminary testing) is not met (L. Wang et al., 2022; Ex6)), or that have been identified as so-called straight-liners (P3). This describes data sets where the participant almost always gives the same answer on the same response scale. All previous types of outliers can be summarized under "error outliers" which can be removed or corrected. However, some outliers might also be authentic answers which provide important or unexpected insights which should not be removed but studied (Aguinis et al., 2013; P4). For an overview of approaches for identification and handling, see Aguinis et al. (2013) and Kim et al. (2019).

Therefore, we propose: Document **(G11)** which data sets are eliminated and why by defining exclusion criteria (P2) and handling strategies to ensure transparency.

8.4 Contributions, Limitations and Future Research

This paper presents a framework consisting of 11 guidelines which describe the survey development process. The guidelines begin after deciding on a theory to be evaluated and include choosing variables, arranging them, ensuring quality, testing the survey, and end before the data is analyzed. A focus is on the information that should be provided when writing the paper. This framework was derived from a structured literature search among the papers of the last ten years in the Basket of Eight journals to ensure high quality in the IS field and validated by three focus groups. Thus, we combine existing knowledge to create a standardized process for survey development. The framework is addressed to beginners who need support in developing a survey, but also more experienced researchers can benefit from a standardized approach, as well as to experts, as open issues regarding the survey development process are identified, and to reviewers in order to give them an overview of the most important aspects of a survey for their feedback. To ensure this, the three focus groups were conducted among experts, proficient, and inexperienced researchers, who all agree that this framework will be helpful in the development process to have an "overview" (Ex6) to "go through the process from front to back" (IP4) and "having the requirements for a survey all at a glance, so that you don't have to gather them from different sources" (P1).

We provide several **contributions for IS research**, especially with regard to quantitative research methodology. First, with our framework, we present a development process that can be applied to the creation of surveys and thus lead to high quality surveys by standardizing them in order to achieve comparability and reproducibility. This is increased at the same time through the focus on the documentation. Thus, we especially enable novice researchers to face the challenge of quantitative data collection and to deliver rigor results. Second, our framework provides a structure that can be used to standardize other methodologies such as experiments or longitudinal surveys. Although DSR is most commonly used to derive design for (IT) artifacts, we adopt it to derive our framework. This is common in research on methodologies (e.g., Nickerson et al., 2009; vom Brocke et al., 2009). Thus, we would like to emphasize with our paper that the DSR approach is very versatile. Third, while our framework focuses primarily on cross-sectional survey development, it can be applied to longitudinal surveys and to some extent to experiments. However, they may not be suited for published statistics because we prioritize the design and data collection phase of survey development and neither cover the process of analyzing the collected data set nor reporting of the results. In addition, we assume that the framework is to some extent transferable to related research areas. The framework for the survey development process can also be utilized in **practice**, as surveys are regularly conducted here as well, which will be of higher quality after applying the guidelines. This could also lead to surveys created in practice being interesting and relevant for research.

Our paper is subject to some **limitations**. As we focus our literature search on high quality papers from the Basket of Eight journals, we exclude a large number of papers that are published at conferences. However, since the aim is to provide the most comprehensive framework possible, we assume that more of the survey development process is documented and reported in the considered papers. In addition, the number of participants in the focus groups is limited, but we tried to include different levels by dividing them according to their knowledge. Finally, our framework cannot guarantee internal validity, even if all guidelines are followed as this also depends on the selection of the theory and the analysis and interpretation of the results. But our framework is not intended to be prescriptive or limiting. Since IS is a very interdisciplinary field, we cannot present all possible issues. Moreover, to some aspects such as the length of the survey, there is no consensus neither in the literature nor among the participants of the focus group. This leads to **future research**. For example, open topics like how to decide for a theory, how to derive hypotheses, the number of participants depending on the sample and the outlet, especially also for pretesting, the optimal length of a survey, or additional methods to prevent biases or ensure validity and reliability during the survey development process should be explored. Also, a comparison between the requirements of a survey and the final implementation and reporting is interesting. We will follow up the paper with a validation of the framework by conducting a sample survey where all guidelines are applied and described in detail.

8.5 Acknowledgements

This research and development project is/was funded by the German Federal Ministry of Education and Research (BMBF) within the "Innovations for Tomorrow's Production, Services, and Work" Program (funding number 02L19C150) and implemented by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication.

9 Discussion and Conclusion

The rapid advancement of AI technologies has led to their increasing significance in both everyday life and various professional contexts. From voice assistants aiding with daily tasks to personalized recommendations in online shops and sophisticated analytical tools in medicine, the potential applications of AI are extensive and continuously expanding (Dwivedi et al., 2023). These developments are primarily driven by key technological innovations such as machine learning and deep learning, which have laid the foundation for today's AI revolution. More recently, the introduction of Generative AI (GenAI), exemplified by systems like ChatGPT, has further transformed the landscape, enabling AI to play a role not only in automation but also in creative processes (OpenAI, 2022; Dwivedi et al., 2023).

AI as a general-purpose technology is here to stay (Berente et al., 2021; Brynjolfsson & McAfee, 2017). AI is not just a technological trend but a transformative force with the potential to revolutionize industries, enhance productivity, and fundamentally alter the way individuals and organizations operate. However, these benefits cannot be fully realized without a comprehensive understanding of the adoption process. The complexity of AI technologies, including issues related to the explainability of AI models (often referred to as the "Black Box" problem) and the requirement for large volumes of high-quality data, poses significant challenges that need to be addressed to facilitate broader adoption (Berente et al., 2021; Polisetty et al., 2024; Pumplun et al., 2019). Additionally, the societal implications of AI, such as ethical concerns and the need for regulatory frameworks, further complicate the adoption (Alzebda & Matar, 2024; Floridi & Cowls, 2019).

Thus, for AI technologies to fully realize their potential, they must first be adopted by users across various levels—social, organizational, and individual. At the social level, cultural attitudes towards technology, the presence of supportive legal and regulatory environments, and the overall societal readiness to embrace AI play critical roles in determining the pace and extent of AI adoption (Floridi & Cowls, 2019; Kumar et al., 2024; Sharma et al., 2022). At the organizational level, factors such as leadership support, strategic alignment, and the availability of technical expertise are crucial for integrating AI into business operations (Pumplun et al., 2019; Radhakrishnan & Chattopadhyay, 2020; Russo, 2024). On the individual level, personal attitudes towards technology, perceived usefulness, ease of use, and trust in AI systems significantly influence whether and how AI technologies are adopted (Henry et al., 2022; Radhakrishnan & Chattopadhyay, 2020; Venkatesh, 2022). Furthermore, only a sound methodological basis for research on AI adoption can provide a deeper understanding of the various dimensions of AI adoption (Recker, 2021).

By examining AI adoption from multiple perspectives—(1) social, (2) organizational, (3) individual, and (4) methodological—this dissertation aims to contribute to the broader discourse on how to harness the transformative potential of AI in a way that maximizes its benefits while mitigating associated risks.

The following sections include the theoretical and practical contributions of this dissertation. Limitations and suggestions for future research arising from the six papers in this dissertation are provided in the respective publications.

9.1 Theoretical Contributions

Paper A addresses RQ1 and explores national culture as an important influencing factor on emerging technologies such as AI in managing organizations, privacy, e-commerce, and especially technology adoption. The structured literature review identified several characteristics that are typical for the measurement of culture in IS literature, as well as a research agenda to guide future research to further topics, extending existing theories, and testing new data analysis methods. With this, Paper A provides an overarching IS-centric overview on the influence of national culture on the adoption of emerging technologies such as AI conceptualized in a concept matrix. It contains comprehensive insights on the culture-related IS theories and diverse research methods are utilized regarding research on the adoption of emerging technologies. Furthermore, we provide a research agenda, which supports the identification of research gaps for future research by combining and merging existing knowledge especially by showing potential research streams regarding the effect of national culture on technologies. Our focus and main findings are that, more diverse theories should be utilized and attention must be given to the importance of measuring national culture at an individual level (e.g., Srite & Karahanna, 2006), especially since collaboration between humans and machines is crucial when using emerging technologies such as AI. New empirical methods can also be used for this purpose, as long as they promise sound and reliable insights. Thus, the social perspective regarding the influence of culture on the adoption of AI is on important factor to consider.

Paper B focuses on the organizational perspective and thus on RQ2. The paper derives drivers and barriers of AI adoption for the two industries financial services and manufacturing. Although some general drivers and barriers to AI adoption have been identified in the last years (e.g., Zöll et al. (2022); Kar et al. (2021); Cubric et al. (2020) and Pumplun et al. (2019)), they cannot explain why AI is utilized in some industries more than others. Thus, resulting from a case study including seven interviews, drivers and barriers of AI adoption were categorized in the TOE framework and included several into soft factors (industry-specific) and hard factors (generalizable). Our results indicate a lack of technological drivers, presumably as those are independent of the industry. In addition, regarding the hard factors (e.g., available data and AI model accuracy), we can support existing literature such as Cubric (2020) regarding AI adoption. In the context of barriers, generalization becomes much more difficult as more soft factors and thus industry-specific barriers are found. Thus, this paper contributes with an overview of the drivers and barriers to AI adoption in financial services and the manufacturing industry and a comparison between those lead to a classification into soft factors (industry-specific) and hard factors (generalizable). These extend existing research on AI adoption (e.g., Zöll et al. (2022); Kar et al. (2021); Cubric et al. (2020) and Pumplun et al. (2019)). Also, initial evidence suggests that these drivers and barriers influence the adoption of AI to a varying degree, which can be important for future AI development or the adoption of other technologies. In addition, due to the combination of Eisenhardt (1989) and Yin (2014) into one methodology, we were able to have a deep insight into the industries to derive the different factors of AI adoption. Furthermore, by conducting interviews, it was possible to draw a picture beyond correlated AI adoption variables and identify dynamic cause-and-effect relationships. Thus, Paper B contributes to the organizational perspective by providing industry-specific factors of AI adoption.

Paper C also relates to RQ2 and the organizational perspective. Focusing on the willingness to pay for ML-based software testing solutions, this paper considers an important factor in the adoption of AI. Although ML-tools can reduce costs, decrease the time required and prevent human errors, some organizations seem to be hesitant in adopting AI. One reason could be the pricing of these tools. Thus, following a two-step process first a structured literature review and subsequently, a Delphi study with experts in the field of ML-based software testing was conducted to determine the required attributes, levels, and prices of such tools. Afterwards, a choice-based conjoint analysis revealed that especially the attribute accuracy is a key element to adopting AI next to the price. With this paper we contribute to the adoption of AI by pointing out economic aspects—especially with the WTP—that are essential when selling these tools to users are considered (Barney et al., 2012). Price is the most important attribute from the users' view on ML-based software testing tools. Although attributes such as accuracy (Briand, 2008), ease of use (Arora et al., 2015; Dejaeger et al., 2013), or explainability (Bouktif et al., 2014; Schieferdecker, 2020) are already mentioned in literature to have an influence on the adoption of AI, we can put a price on it. For example, accuracy is the second most essential attribute for potential users after the price itself. The participants are willing to pay up to €120 per license per month more for an increase in accuracy from 90% to 99%. In addition, integration and ease of use are also perceived by the participants as being relevant and they both receive a marginal willingness to pay of around €20. Thus, we not only determined new attributes, as well as their levels and prices as part of our Delphi study, but also examined their value for users. Thus, with Paper C we can make an important contribution regarding the purchase and thus the adoption of AI tools in organizations.

Paper D focuses on the individual perspective (RQ 3) by examining the influence of GenAI on the performance and meaningfulness of having ChatGPT as an assistance for coding. By conducting an online experiment with two types of tasks—coding and debugging, we were able to show that consistent with previous studies on human-AI collaboration (Boyacı et al., 2023; Fügener et al., 2021) participants performed higher in their tasks when they had assistance from ChatGPT. This was especially due to an indirect effect via task difficulty, thus ChatGPT can simplify tasks. Regarding meaningfulness we only found a significant decrease for the coding task, mediated by task difficulty. These findings are in line with research in psychology that has indicated that increased effort can sometimes make tasks feel more meaningful for individuals (Campbell et al., 2022; Gielnik et al., 2015; Inzlicht & Campbell, 2022; Mortimer, 2023). Finally, our findings are depending on the differences of the two programming tasks, which indicates that AI adoption is also depending on the types of tasks. With these findings, this paper offers several contributions regarding AI adoption. Due to the increase in task performance and thus a successful human-AI collaboration, the adoption of AI can be improved. However, as the assistance of ChatGPT decrease perceived meaningfulness, we combined IS literature and a known phenomenon from psychology (see Bailey et al., 2019). This is especially relevant as Gen Z, for example, rate meaningfulness as one of the main factors to take a new job (McKinsey, 2022). Depending on the task, we observed a decrease in perceived meaningfulness with ChatGPT assistance. It is possible that AI reduces the effort we have to invest in our work, leading to a perception of our work as less meaningful. Since these results are mixed, we can only provide an initial assessment of the influence of perceived meaningfulness on AI adoption. Lastly, the adoption of AI is also depending on the type of task. Replacing creative or challenging tasks might lead to less willingness to adopt AI compared to replacing or collaboratively solving relatively easy, boring, or repetitive tasks with (generative) AI. Such insights can pave the way for more holistic strategies in AI adoption, ensuring that technological advancements are harmoniously aligned with human well-being and purpose.

Paper E also addresses RQ3 and examines the individual perspective of AI adoption. Utilizing an online experiment, we examined if the IKEA effect, a bias that states that humans perceive the value of outcomes higher, if they poured more effort into it (Mochon et al., 2012; Norton et al., 2012), can also be found when using GenAI. The IKEA effect can influence the AI adoption as human-AI interaction can result in higher WTP, appreciation and value of the final outcome compared to the automatization of processes. Therefore, our participants had to solve a text and an image task once in co-creation with a GenAI tool and once received the GenAI solution immediately. Our findings show the presence of the IKEA effect for the image generation task, where the effort manipulation was successful. Thus, this paper contributes to the individual AI adoption by showing that although AI has unique characteristics such as being a black box (Bauer, Hinz, van der Aalst, et al., 2021; Dwivedi et al., 2023), it can appear if individuals perceive a significant difference in the achievement and contribution to an output. In addition, with the ability of GenAI to create new content instead of solely making decisions, we show with our findings that human-AI collaboration needs to be rethought as they now participate in the outcome and co-develop the result with their input. Thus, it is important for the adoption of AI that humans and AI collaborate, even though humans might not fully understand the functionality of the AI. Therefore, the IKEA effect should be leveraged and could be integrated artificially into the collaboration of humans and machines. This could increase the adoption rate of AI.

Finally, Paper F uses a design science research approach consisting of a structured literature review and focus groups to derive a framework consisting of 11 guidelines which describe the survey development process and thus addressing RQ4 to provide a sound research methodology for research on the adoption of AI. The guidelines focus on the information that should be provided when writing a paper which utilized a cross-sectional survey and begins after deciding on a theory to be evaluated and includes choosing variables, arranging them, ensuring quality, testing the survey, and ends before the data is analyzed. With Paper F we provide several contributions by providing a framework that is addressed to beginners who need support in developing a survey, but also more experienced researchers can benefit from a standardized approach, as well as to experts, as open issues regarding the survey development process are identified, and to reviewers in order to give them an overview of the most important aspects of a survey for their feedback. Thus, with our guidelines researchers can develop cross-sectional surveys on topics regarding the AI adoption. But our procedure can also be applied to standardize other methodologies for AI adoption research as a sound methodology is an important factor in IS research.

In summary, the six papers in this dissertation present new insights into AI adoption. By examining AI adoption from multiple perspectives—social, organizational, individual, and methodological—this research aims to contribute to the broader discourse of influencing factors on AI adoption.

9.2 Practical Contributions

In addition to the theoretical contributions, the six papers included in this dissertation also have practical contributions, especially for organizations, but also for individuals and the society. These practical contributions contribute to successful AI adoption.

Paper A addresses RQ1 and gives insights in the influence of culture on emerging technologies such as AI by conducting a structured literature review, thus a social perspective. Knowing the influence of culture, organizations operating across national borders and serving customers from diverse cultural backgrounds can foster teamwork, retain and attract international customers and most importantly successfully adopt emerging technologies. Thus, organizations can benefit from the presented knowledge to perform their own analyses with respect to introducing emerging technologies. Also, we point out that organizations should not ignore cultural differences when applying e.g. management formulas from their cultural context. Especially when adopting technologies such as AI, there is no one-size fits all approach. We support this by guiding organizations with the identified and collected literature which include previously examined cultural influences and derived recommendations. Thus, including a social perspective regarding national culture can improve the adoption of AI.

Paper B focuses on RQ2 and has an organizational perspective, thus especially include implications for industry stakeholders, business decision-makers, and AI executives. By deriving industry-specific drivers and barriers for AI adoption for financial service and manufacturing from seven interviews as a multiple-case study, the factors provide cross-company knowledge, enabling market and technology transparency. They provide information to understand the current industry situation and the underlying context. Thus, this can serve both as a starting point for market analyses and technological investigation of AI adoption. Stakeholders in those industries can utilize these drivers and barriers for their AI adoption. Also, AI specialists or consultants can benefit from understanding their customers' needs, challenges, and motivating influences behind their AI adoption journeys. Thus, organizations can profit from these factors for their AI adoption.

Paper C also relates to RQ2 and has an organizational perspective on AI adoption. By conducting a literature review, a Delphi study and a conjoint analysis, we identified factors and examined the willingness to pay for AI regarding ML-based software testing tools. The findings not only provide insights to the organizations adopting AI, but also the developers and suppliers of AI. We especially recommend to develop more AI tools, especially as many companies have a relative high budget available. The pricing strategy here could be gaining market shares via initially low prices and bounding users by lock-in effects and increasing the price over time. In addition, the willingness to pay of different user types can be skimmed off by offering different product combinations. Furthermore, segmentation of the customers could increase the number of licenses sold, e.g., by making the integration feature optional, as older participants revealed a negative WTP. Thus, knowing the WTP and the desired features as a developer of AI can increase the adoption of these tools.

Paper D addresses the individual perspective, thus RQ3, but still has important implications for organizations. The results from an online experiment with experienced coders show that GenAI assistance leads to higher performance but lower meaningfulness due to facilitation of the tasks. Thus, organizations should ensure that employees' perceived meaningfulness remains intact when working with GenAI as perceived meaningfulness leads to more motivation, which, in turn, results in higher company performance (Allan, 2017). Therefore, it is important to consider which tasks can be performed collaboratively with AI, such as replacing routine tasks while jointly working on complex tasks or letting humans perform creative tasks. Also, job crafting could be a solution to generate more meaningfulness. In addition, our findings suggest that organizations should promote the use of GenAI due to the increase in task performance, as higher individual performance will lead to higher organizational performance. However, while allowing the use of such tools, organizations should evaluate which tools are most beneficial, especially also regarding privacy and security of the organization's data. Therefore, considering the psychological influences of the usage of AI can increase the adoption in organizations and by individuals.

Paper E also focuses on the individual perspective and thus RQ3 by conducting an online experiment to assess the influence of the IKEA effect when using GenAI. Our results can be utilized by developers of AI tools and organizations adopting AI. As we found a higher WTP and behavioral intention to use for GenAI with human participation and thus the IKEA effect, developers can exploit this by adapting strategies of customizing instead of exclusively selling automation solutions to customers. This is especially useful for high-stakes environments where humans are likely to remain the final decision makers in the foreseeable future. But also, organizations adopting AI can utilize the IKEA effect by implementing AI solutions that require collaboration between humans and AI. This can result in employees being less afraid of losing their jobs, resulting in higher acceptance of those solutions. In addition, the collaboration between humans and machines can increase the perceived value of the end product which increases the behavioral intention to use and thus higher acceptance and adoption rates. Thus, organizations should utilize biases in human behavior to increase the adoption of AI and recommend potential users of AI to give it a try and discover new ways of facilitating work.

Finally, Paper F addresses the methodological perspective and therefore RQ4. Although having sound methodologies is mainly important for research, the framework for the survey development process derived from a DRS approach utilizing a structured literature review and focus groups can also be utilized in practice. Organizations also conduct surveys regularly, which will be of higher quality after applying the guidelines. Utilizing the guidelines and ensure a sound research approach might even result in surveys being interesting and relevant for research.

Overall, the papers included in this dissertation provide several practical contributions to support organizations, but also society and individuals in their AI adoption. By identifying and analyzing different factors that influence the adoption of AI, the power of AI can be exploited.

9.3 Concluding Remarks

As AI has become an integral part of everyday life (Berente et al., 2021; Brynjolfsson & Mitchell, 2017; Russell & Norvig, 2021), this dissertation explores the multifaceted process of AI adoption. Despite the rapid advancements in AI, as evidenced by the emergence of GenAI and tools like ChatGPT, the widespread adoption of AI remains low. Thus, this dissertation aimed to investigate the factors that influence AI adoption from a social, organizational, and individual perspective, as well as to provide a sound methodological approach. To answer the research questions, this dissertation includes six papers, one structured literature review, one case study, one conjoint analysis, two online experiments and a design science research approach investigating different factors affecting the AI adoption.

Although the included papers already present specific directives for future research, the following suggestions for future research derive from the four perspectives of this dissertation. First, regarding the *social perspective* (RQ1) we showed that culture is an important influence on the adoption of emerging technologies such as AI. Future research can build on these findings and expand research regarding newer developments of AI such as GenAI. While Paper A focuses on the cultural influence, other important factors of a society such as their norms and values, which leads to their regulations and laws, but also their technological development can play a role in AI adoption. Thus, expanding the social perspective by including more factors can help societies, organizations and individuals in their AI adoption.

Second, from the *organizational perspective* (RQ2) this dissertation included Paper B and C. Paper B derived industry-specific drivers and barriers of AI adoption. This can especially be expanded for other industries such as healthcare and for newer technological developments such as GenAI as drivers and barriers can be influenced. In addition, quantitative research on the intensity of these factors and on how to overcome these barriers and exploit drivers can assist organizations in AI adoption. Also, Paper C, which examines the WTP for AI tools, can be a starting point for future research. Additionally, to the attributes of accuracy, integration and ease of use we identified as the most important, more characteristics of AI tools can be considered. As this paper explores one specific type of AI—ML-based software testing tools, this can be broadened by measuring the WTP for other types of AI, which can help to generalize these results. Furthermore, not only nations have their respective culture, which was followed from Paper A, but also organizations, which might influence their adoption rate. Thus, research regarding organizational adoption of AI can consider culture as an influencing factor. Overall, many more specific factors for AI adoption in organizations can be part of future research to assist the adoption process.

Third, considering the *individual perspective* (RQ3) Paper D and E examined psychological factors as influencing factors for the AI adoption. While Paper D examines the influence of GenAI assistance on the performance and perceived meaningfulness, many more factors are important to individuals to use these tools. Future research could therefore consider factors such as attitude towards technologies or experience with AI. But also, the usability and explainability of AI tools is important when adopting AI. Here, theories such as the "task-technology-fit" theory (see Sturm & Peters, 2020) can be utilized by future research. In addition, while national culture is important, a key finding of Paper A was that culture should be measured on an individual level, including characteristics and psychological factors into the adoption process. In an organizational context this could be the ease of use or the WTP from Paper C, which are important factors not only for organizations to adopt AI, but also individuals and needs further research. Paper E finds that the IKEA effect, a bias where outputs generated with more effort are considered to have more value, also leads to future research. Showing that biases that are so far only considered for physical products or interactions can be transferred to human-AI collaboration, opens new possibilities regarding future research. These psychological factors either can be prevented to increase AI adoption or exploited. Future research should examine not only which biases can have an effect, but also how to utilize or block them. Thus, individual factors and biases are important when considering AI adoption.

Finally, Paper F adopts the *methodological perspective* (RQ4) and provides a sound research approach to support AI adoption research. Future research can on the one hand use this to find influencing factors on AI adoption and analyze their importance. On the other hand, this paper can be used as a template to standardize and facilitate other research approaches.

Overall, society, organizations and individuals face different influencing factors when adopting AI. In this dissertation, I aim to provide some of them to guide underlying research regarding the adoption of AI. Future research can build on these identified factors and various stakeholders can use the findings to better adopt AI. Thus, this dissertation contributes to a deeper understanding of the conditions necessary for successful AI adoption. By analyzing the various dimensions of AI integration, it provides valuable insights into how AI can be effectively adopted by individuals, incorporated into organizations and society, ultimately maximizing its potential by utilizing a sound research methodology.

References

Abbey, J. D., & Meloy, M. G. (2017). Attention by Design: Using Attention Checks to Detect Inattentive Respondents and Improve Data Quality. *Journal of Operations Management*, *53–56*, 63–70.

Abdalla, A. A., Bhat, M. A., Tiwari, C. K., Khan, S. T., & Wedajo, A. D. (2024). Exploring ChatGPT adoption among business and management students through the lens of diffusion of Innovation Theory. *Computers and Education: Artificial Intelligence*, *7*(June), 100257. https://doi.org/10.1016/j.caeai.2024.100257

Abel-Karim, B. M., Pfeuffer, N., Rohde, G., & Hinz, O. (2020). How and What Can Humans Learn from Being in the Loop? Invoking Contradiction Learning as a Measure to Make Humans Smarter. *KI - Künstliche Intelligenz*, *34*, 199–207.

Abhari, K., Davidson, E., & Xiao, B. (2022). Inventing Together: The Role of Actor Goals and Platform Affordances in Open Innovation. *Journal of the Association for Information Systems*, *23*(1), 264–302.

Adidas. (2023). *Personalisable*. https://www.adidas.de/personalisable

Afrouzi, O. A. (2021). Humanitarian Behavior Across High- / Low- Context Cultures: A Comparative Analysis Between Switzerland and Colombia. *Journal of International Humanitarian Action*, *6*(2). https://doi.org/https://doi.org/10.1186/s41018-020-00088-y

Agarwal, R., Ahuja, M., Carter, P., & Gans, M. (1998). Early and late adopters of IT innovations: extensions to innovation diffusion theory. *Proceedings of the DIGIT Conference*, *August 2015*, 1–18. http://disc-nt.cba.uh.edu/chin/digit98/panel2.pdf

Agrawal, A., Gans, J. S., & Goldfarb, A. (2024). Artificial intelligence adoption and system-wide change. *Journal of Economics and Management Strategy*, *33*(2), 327–337. https://doi.org/10.1111/jems.12521

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). *Best-Practice Recommendations for Defining, Identifying, and Handling Outliers*. *16*(2), 270–301.

Ahmed, B. S., & Zamli, K. Z. (2011). A Variable Strength Interaction Test Suites Generation Strategy Using Particle Swarm Optimization. *Journal of Systems and Software*, *84*(12), 2171–2185.

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211.

Albani, A., Domigall, Y., & Winter, R. (2017). Implications of Customer Value Perceptions for the Design of Electricity Efficiency Services in Times of Smart Metering. *Information Systems and E-Business Management*, *15*(4), 825–844.

Albaum, G. (1997). The Likert Scale Revisited: An Alternate Version. *Journal of the Market Research Society*, *39*(2), 331–348.

Ali, K., Alzaidi, M., Al-Fraihat, D., & Elamir, A. M. (2023). Artificial Intelligence: Benefits, Application, Ethical Issues, and Organizational Responses. In A. K. Nagar, D. Singh Jat, D. K. Mishra, & A. Joshi (Eds.), *Intelligent Sustainable Systems. Lecture Notes in Networks and Systems, vol 578*. Springer.

Allan, B. A. (2017). Task Significance and Meaningful Work: A Longitudinal Study. *Journal of Vocational Behavior*, *102*, 174–182.

Alzebda, S., & Matar, M. A. (2024). Factors Affecting Citizen Intention Toward AI Acceptance and Adoption: The Moderating Role of Government Regulations. *Competitiveness Review: An International Business Journal*. https://doi.org/https://doi.org/10.1108/CR-06-2023-0144

Armstrong, J. S., & Overton, T. S. (1977). Estimating Nonresponse Bias in Mail Surveys. *Journal of*

Marketing Research, *14*(3), 396–402.

Arora, I., Tetarwal, V., & Saha, A. (2015). Open Issues in Software Defect Prediction. *Procedia Computer Science*, *46*, 906–912.

Atinc, G., Simmering, M. J., & Kroll, M. J. (2012). Control Variable Use and Reporting in Macro and Micro Management Research. *Organizational Research Methods*, *15*(1), 57–74.

Awad, E., Dsouza, S., & Kim, R. (2018). The Moral Machine Experiment. *Nature*, *563*, 59–64.

Baggen, R., Correia, J. P., Schill, K., & Visser, J. (2012). Standardized Code Quality Benchmarking for Improving Software Maintainability. *Software Quality Journal*, *20*(2), 287–307. https://doi.org/10.1007/s11219-011-9144-9

Bailey, C., Lips-Wiersma, M., Madden, A., Yeoman, R., Thompson, M., & Chalofsky, N. (2019). The Five Paradoxes of Meaningful Work: Introduction to the Special Issue 'Meaningful Work: Prospects for the 21st Century. *Journal of Management Studies*, *56*(3), 481–499.

Baker, J., Parasuraman, A., Grewal, D., & Voss, G. B. (2002). The Influence of Multiple Store Environment Cues on Perceived Merchandise Value and Patronage Intentions. *Journal of Marketing*, *66*(2), 120–141.

Balakrishnan, J., Dwivedi, Y. K., Hughes, L., & Boy, F. (2021). Enablers and Inhibitors of AI - Powered Voice Assistants : A Dual - Factor Approach by Integrating the Status Quo Bias and Technology Acceptance Model. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-021-10203-y

Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, *33*(1), 1–17. https://doi.org/10.1007/s12525-023-00680-1

Bankins, S., & Formosa, P. (2020). When AI Meets PC: Exploring the Implications of Workplace Social Robots and a Human-Robot Psychological Contract. *European Journal of Work and Organizational Psychology*, *29*(2), 215–229. https://doi.org/10.1080/1359432X.2019.1620328

Bankins, S., & Formosa, P. (2023). The Ethical Implications of Artificial Intelligence (AI) For Meaningful Work. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-023-05339-7

Bao, Y., Hilary, G., & Ke, B. (2022). Artificial Intelligence and Fraud Detection. In V. Babich, J. R. Birge, & G. Hilary (Eds.), *Innovative Technology at the Interface of Finance and Operations. Springer Series in Supply Chain Management, vol 11*. Springer, Cham.

Barkhi, R., & Kao, Y.-C. (2011). Psychological Climate and Decision-Making Performance in a GDSS Context. *Information \& Management*, *48*(4–5), 125–134. https://doi.org/10.1016/j.im.2011.02.003

Barney, S., Petersen, K., Svahnberg, M., Aurum, A., & Barney, H. (2012). Software Quality Trade-Offs: A Systematic Map Sebastian. *Information and Software Technology*, *54*(7), 651–662.

Baron, R., & Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personalitiy and Social Psychology*, *51*(6), 1173–1182.

Barr, E. T., Harman, M., McMinn, P., Shahbaz, M., & Yoo, S. (2015). The Oracle Problem in Software Testing: A Survey. *IEEE Transactions on Software Engineering*, *41*(5), 507–525.

Baskerville, R. L., Kaul, M., & Storey, V. C. (2015). Genres of Inquiry in Design-Science Research. *MIS Quarterly*, *39*(3), 541–564.

Bauer, K., Hinz, O., Aalst, W. Van Der, & Weinhardt, C. (2021). Expl(AI)n it to me–explainable AI and information systems research. *Business & Information Systems Engineering*, *63*(2), 79–82.

Bauer, K., Hinz, O., van der Aalst, W., & Weinhardt, C. (2021). Expl(AI)n it to me–explainable AI and information systems research. *Business & Information Systems Engineering*, *63*(2), 79–82.

Bedué, P., & Fritzsche, A. (2022). Can We Trust AI? An Empirical Investigation of Trust Requirements and Guide to Successful AI Adoption. *Journal of Enterprise Information Management*, *35*(2), 530–549. https://doi.org/https://doi.org/10.1108/JEIM-06-2020-0233

Beens, R. E. G. (2020). *The Privacy Mindset Of The EU Vs. The US*. Forbes. https://www.forbes.com/sites/forbestechcouncil/2020/07/29/the-privacy-mindset-of-the-eu-vs-the-us/?sh=233728087d01

Bellman, S., Johnson, E. J., Kobrin, S. J., & Lohse, G. L. (2004). *International Differences in Information Privacy Concerns : A Global Survey of Consumers*. 313–324.

Benbya, H., Pachidi, S., & Jarvenpaa, S. L. (2021). Special issue editorial: Artificial intelligence in organizations: Implications for information systems research. *Journal of the Association for Information Systems*, *22*(2), 281–303. https://doi.org/10.17705/1jais.00662

Benlian, A. (2015). IT Feature Use over Time and its Impact on Individual Task Performance. *Journal of the Association for Information Systems*, *16*(3), 144–173. https://doi.org/10.17705/1jais.00391

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing Artificial Intelligence. *MIS Quartley*, *45*(3), 1433–1450.

Bergkvist, L., & Rossiter, J. R. (2007). The Predictive Validity of Multiple-Item Versus Single-Item Measures of the Same Constructs. *Journal of Marketing Research*, *44*(2), 175–184.

Bernerth, J. B., & Aguinis, H. (2016). A Critical Review and Best-Practice Recommendations for Control Variable Usage. *Personel Psychology*, *69*, 229–283.

Bhalerao, K., Kumar, A., Kumar, A., & Purvi, P. (2022). A Study of Barriers and Benefits of Artificial Intelligence Adoption in Small and Medium Enterprises. *Academy of Marketing Studies Journal*, *26*(S1), 1–6. https://www.researchgate.net/publication/360912025_A_STUDY_OF_BARRIERS_AND_BENEFITS_OF_ARTIFICIAL_INTELLIGENCE_ADOPTION_IN_SMALL_AND_MEDIUM_ENTERPRISE

Bhattacherjee, A. (1998). Management of Emerging Technologies : Experiences and Lessons Learned at US West. *Information & Management*, *33*, 263–272.

Böhm, S., Adam, F., & Farrell, W. C. (2015). Impact of the Mobile Operating System on Smartphone Buying Decisions: A Conjoint-Based Empirical Analysis. *International Conference on Mobile Web and Information Systems*, 198–210.

Bolton, R. N. (1993). Pretesting Questionaires: Content Analyses of Respondent's Concurrent Verbal Protocols. *Marketing Science*, *12*(3), 280–303.

Boogerd, C., & Moonen, L. (2008). Assessing the Value of Coding Standards: An Empirical Study. *2008 IEEE International Conference on Software Maintenance*, 277–286. https://doi.org/10.1109/ICSM.2008.4658076

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the Criterion Domain to Include Elements of Contextual Performance. In *Personnel selection in organizations*. Jossey-Bass.

Borman, W. C., & Motowidlo, S. J. (1997). Task Performance and Contextual Performance: The Meaning for Personnel Selection Research. *Human Performance*, *10*(2), 99–109. https://doi.org/10.1207/s15327043hup1002_3

Bouktif, S., Sahraoui, H., & Ahmed, F. (2014). Predicting Stability of Open-Source Software Systems Using Combination of Bayesian Classifier. *ACM Transactions on Management*

Information Systems (TMIS), *5*(1), 1–26.

Boyacı, T., Canyakmaz, C., & Véricourt, F. De. (2023). Human and Machine: The Impact of Machine Input on Decision Making Under Cognitive Limitations. *Management Science*. https://doi.org/10.1287/mnsc.2023.4744

Brady, M. K., & Robertson, C. J. (1999). An Exploratory Study of Service Value in the USA and Ecuador. *International Journal of Service Industry Management*, *10*(5), 469–486.

Bredehorst, J., Krautter, K., Meuris, J., & Jachimowicz, J. M. (2023). The Challenge of Maintaining Passion for Work over Time: A Daily Perspective on Passion and Emotional Exhaustion. *Organization Science*, orsc.2023.1673. https://doi.org/10.1287/orsc.2023.1673

Breidert, C., Hahsler, M., & Reutterer, T. (2006). A Review of Methods for Measuring Willingness-To-Pay. *Innovative Marketing*, *2*(4), 8–32.

Bremser, C. (2018). Starting Points for Big Data Adoption. *Proceedings of the 26th European Conference on Information Systems (ECIS)*.

Briand, L. C. (2008). Novel Applications of Machine Learning in Software Testing. *The Eighth International Conference on Quality Software*, 3–10.

Brislin, R. W. (1970). Back-Translation for Cross-Cultural Research. *Journal of Cross-Cultural Psychology*, *1*(3), 185–216.

Bryant, K. (2023). *How AI Will Impact The Next Generation Workforce*. https://www.forbes.com/sites/kalinabryant/2023/05/31/how-ai-will-impact-the-next-generation-workforce/

Brynjolfsson, E., & McAfee, A. (2017). The Business of Artificial Intelligence: What It Can - And Cannot - Do for Your Organization. *Harvard Business Review*.

Brynjolfsson, Erik, & Mitchell, T. (2017). What Can Machine Learning Do? Workforce Implications. *Science*, *358*(6370), 1530–1534.

Buchanan, B. G., & Wright, D. (2021). The Impact of Machine Learning on UK Financial Services. *Oxford Review of Economic Policy*, 537–563.

Build-A-Bear. (2023). *The-Bear-Builder*. https://www.buildabear.co.uk/on/demandware.store/Sites-buildabear-uk-Site/en_GB/BearBuilder-Show?step=chooseFriends

Campbell, A. V., Chung, J. M. H., & Inzlicht, M. (2022). Meaningfulness of Effort: Deriving Purpose From Really Trying. *PsyArXiv*.

Capgemini. (2023). *Why Consumers Love Generative AI*. https://prod.ucwe.capgemini.com/wp-content/uploads/2023/06/GENERATIVE-%0AAI_Final_WEB_060723.pdf

Carlson, K. D., & Wu, J. (2012). The Illusion of Statistical Control: Control Variable Practice in Management. *Organizational Research Methods*, *15*(3), 413–435.

Catal, C., & Diri, B. (2009). Investigating the Effect of Dataset Size, Metrics Sets, and Feature Selection Techniques on Software Fault Prediction Problem. *Information Sciences*, *179*(8), 1040–1058.

Cech, E. (2021). *The Trouble With Passion: How Searching for Fulfillment at Work Fosters Inequality*. Univ of California Press.

Chang, Y. W., Hsu, P. Y., Shiau, W. L., & Tsai, C. C. (2014). Knowledge Sharing Intention in the United States and China: A Cross-Cultural Study. *European Journal of Information Systems*, *24*, 262–277.

Che, T., Peng, Z., Lai, F., & Luo, X. (2022). Online Prejudice and Barriers to Digital Innovation: Empirical Investigations of Chinese Consumers. *Information Systems Journal*, *32*(3), 630–652.

Chen, E., Huang, R., Chen, H.-S., Tseng, Y.-H., & Li, L.-Y. (2023). GPTutor: A ChatGPT-Powered Programming Tool for Code Explanation. In N. Wang, G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda, & O. C. Santos (Eds.), *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky* (Vol. 1831, pp. 321–327). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36336-8_50

Chen, J., Zhou, W., & Frankwick, G. L. (2023). Firm AI Adoption Intensity and Marketing Performance. *Journal of Computer Information Systems*, *00*(00), 1–18. https://doi.org/10.1080/08874417.2023.2277751

Chen, Yan, & Zahedi, F. M. (2016). Individuals' Internet Security Perceptions and Behaviors: Polycontextual Contrasts Between the United States and China. *MIS Quarterly*, *40*(1), 205–222.

Chen, Yi, Wang, R., Jiang, H., Shi, S., & Xu, R. (2023). Exploring the Use of Large Language Models for Reference-Free Text Quality Evaluation: A Preliminary Empirical Study. *ArXiv Preprint ArXiv:2304.00723*.

Chin, W. W., Johnson, N., & Schwarz, A. (2008). A Fast Form Approach to Measuring Technology Acceptance and Other Constructs. *MIS Quarterly*, *32*(4), 687–703.

Chin, W. W., Marcolin, B. L., & Newsted, P. R. (1996). A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results From a Monte Carlo Simulation Study and Voice Mail Emotion/Adoption Study. *ICIS 1996 Proceedings*, *14*(2), 21–41.

Chu, X., Luo, X. (Robert), & Chen, Y. (2019). A Systematic Review on Cross-Cultural Information Systems Research: Evidence From the Last Decade. *Information and Management*, *56*(3), 403–417.

Chuma, E. L., & De Oliveira, G. G. (2023). Generative AI for Business Decision-Making: A Case of ChatGPT. *Management Science and Business Decisions*, *3*(1), 5–11. https://doi.org/10.52812/msbd.63

Churchill, G. A. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research*, *16*(1), 64.

Cichy, P., Salge, T. O., & Kohli, R. (2021). Privacy Concerns and Data Sharing in the Internet of Things: Mixed Methods Evidence from Connected Cars. *MIS Quarterly*, *45*(4), 1863–1892.

Ciolkowski, M., Laitenberger, O., Vegas, S., & Biffl, S. (2003). Practical Experiences in the Design and Conduct of Surveys in Empirical Software Engineering. In *Empirical Methods and Studies in Software Engineering* (pp. 104–128).

Compeau, D., Correia, J., & Thatcher, J. B. (2022). When Constructs Become Obsolete: A Systematic Approach to Evaluating and Updating Constructs for Information Systems Research. *MIS Quarterly*, *46*(2), 679–711.

Conboy, K., Fitzgerald, G., & Mathiassen, L. (2012). Qualitative methods research in information systems: motivations, themes, and contributions. *European Journal of Information Systems*, *21*(2), 113–118. https://doi.org/10.1057/ejis.2011.57

Conger, S., Pratt, J. H., & Loch, K. D. (2013). Personal Information Privacy and Emerging Technologies. *Information Systems Journal*, *23*(5), 401–417.

Conjointly. (2022). *Survey Platform With Easy-To-Use Advanced Tools and Expert Support*. https://conjointly.com/

Cooper, H. M. (1988). Organizing Knowledge Syntheses: A Taxonomy of Literature Reviews. *Knowledge in Society*, *1*(1), 104–126.

Cram, W. A., Wiener, M., Tarafdar, M., & Benlian, A. (2022). Examining the Impact of Algorithmic

Control on Uber Drivers' Technostress. *Journal of Management Information Systems*, *39*(2), 426–453.

Craney, T. A., & Surles, J. G. (2002). Model-Dependent Variance Inflation Factor Cutoff Values. *Quality Engineering*, *14*(3), 391–403.

Cubric, M. (2020). Drivers, Barriers and Social Considerations for AI Adoption in Business and Management: A tertiary study. *Technology in Society*, *62*, 1–12.

Cyr, D. (2008). Modeling Web Site Design Across Cultures: Relationships to Trust, Satisfaction, and E-loyalty. *Journal of Management Information Systems*, *24*(4), 47–72.

Cyr, D., Head, M., Larios, H., & Pan, B. (2009). Exploring Human Images in Website Design: A Multi-Method Approach. *MIS Quarterly*, *33*(3), 539–566.

Dadkhah, M., Araban, S., & Paydar, S. (2020). A Systematic Literature Review on Semantic Web Enabled Software Testing. *Journal of Systems and Software*, *162*, 110485.

Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and Machine Learning Models in Credit Scoring: A systematic literature survey. *Applied Soft Computing*, *91*, 1–21.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 319–340.

Davison, R. M., Martinsons, M. G., Ou, C. X. J., Murata, K., Drummond, D., Li, Y., & Lo, H. W. H. (2009). The Ethics of It Professionals in Japan and China. *Journal of the Association for Information Systems*, *10*(11), 834–859.

de Propris, L., & Bailey, D. (Eds.). (2020). *Regions and cities. Industry 4.0 and Regional Transformation*. Routledge.

DeepL. (2023). *Home*. https://www.deepl.com/de/translator

Dejaeger, K., Verbraken, T., & Baesens, B. (2013). Toward Comprehensible Software Fault Prediction Models Using Bayesian Network Classifiers. *IEEE Transactions on Software Engineering*, *39*(2), 237–257.

DePietro, R., Wiarda, E., & Fleischer, M. (1990). The Context for Change: Organization, Technology and Environment. In L. G. Tornatzky & M. Fleischer (Eds.), *The Processes of Technological Innovation,* (pp. 151-). Lexington Books.

Dietvorst, B. J., & Bharti, S. (2020). People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error. *Psychological Science*, *31*(10), 1302–1314. https://doi.org/10.1177/0956797620948841

Dinev, T., Bellotto, M., Hart, P., Russo, V., Serra, I., & Colautti, C. (2006). Privacy Calculus Model in E-commerce - A Study of Italy and the United States. *European Journal of Information Systems*, *15*, 389–402.

Dinev, T., Goo, J., Hu, Q., & Nam, K. (2009). User Behaviour Towards Protective Information Technologies: The Role of National Cultural Differences. *Information Systems Journal*, *19*(4), 391–412.

Dinev, T., Xu, H., Smith, J. H., & Hart, P. (2013). Information Privacy and Correlates: An Empirical Attempt to Bridge and Distinguish Privacy-related Concepts. *European Journal of Information Systems*, *22*(3), 295–316.

Ding, H., Gao, R. X., Isaksson, A. J., Landers, R. G., Parisini, T., & Yuan, Y. (2020). State of AI-Based Monitoring in Smart Manufacturing and Introduction to Focused Section. *IEEE/ASME Transactions on Mechatronics*, *25*(5), 2143–2154.

Dohle, S., Rall, S., & Siegrist, M. (2014). I Cooked It Myself: Preparing Food Increases Liking and Consumption. *Food Quality and Preference*, *33*, 14–16. https://doi.org/10.1016/J.FOODQUAL.2013.11.001

Dohle, S., Rall, S., & Siegrist, M. (2016). Does Self-Prepared Food Taste Better? Effects of Food Preparation on Liking. *Health Psychology*, *35*(5), 500–508. https://doi.org/10.1037/HEA0000315

Drury, D. H., & Farhoomand, A. (1997). Improving Management Information Systems Research: Question Order Effects in Surveys. *Information Systems Journal*, *7*, 241–251.

Durelli, V. H. S., Durelli, R. S., Borges, S. S., Endo, A. T., Eler, M. M., Dias, D. R. C., & Guimarães, M. P. (2019). Machine Learning Applied to Software Testing: A Systematic Mapping Study. *IEEE Transactions on Reliability*, *68*(3), 1189–1212.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Louise, E., Jeyaraj, A., Kumar, A., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Ahmad, M., Al-busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion Paper: "So What if ChatGPT Wrote It ?" Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy. *International Journal of Information Management*, *71*.

Earley, P. C. (1997). *Face, Harmony, and Social Structure: An Analysis of Organizational Behavior Across Cultures*. Oxford University Press on Demand.

Ebbers, F., Zibuschka, J., Zimmermann, C., & Hinz, O. (2021). User Preferences for Privacy Features in Digital Assistants. *Electronic Markets*, *31*(2), 411–426.

Ebert, C. (2024). Generative AI for Software Practitioners. *IEEE Software*, *40*, 30–38. https://doi.org/10.1109/MS.2023.3265877

Ebert, C., Bajaj, D., & Weyrich, M. (2022). Testing Software Systems. *IEEE Software*, *39*(4), 8–17.

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review*, *14*, 532–550.

Eisenhardt, Kathleen M. (1989). Building Theories from Case Study Research. *Academy of Management Review*, *14*(4), 532–550.

Ellenrieder, S., Kallina, E. M., Pumplun, L., Gawlitza, J. F., Ziegelmayer, S., & Buxmann, P. (2023). Promoting Learning Through Explainable Artificial Intelligence : An Experimental Study in Radiology. *ICIS 2023 Proceedings*, 1–17.

Elmuti, D., Grunewald, J., & Abebe, D. (2010). Consequences of Outsourcing Strategies on Employee Quality of Work Life, Attitudes, and Performance. *Journal of Business Strategies*, *27*(2), 177–203.

Enoiu, E., Tukseferi, G., & Feldt, R. (2020). Towards a Model of Testers' Cognitive Processes: Software Testing as a Problem Solving Approach. *2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, 272–279.

Eurostat. (2023). *Artificial Intelligence by Size Class of Enterprise*. https://ec.europa.eu/eurostat/databrowser/view/isoc_eb_ai/default/table?lang=en

Fang, X., Benamati, J. "Skip," & Lederer, A. L. (2011). Coping With Rapid Information Technology Change in Different National Cultures. *European Journal of Information Systems*, *20*, 560–573.

Farhi, F., Jeljeli, R., Zahre, A., Saudani, S., & Feguiri, L. (2023). Factors Behind Virtual Assistance Usage Among iPhone Users Theory of Reasoned Action. *International Journal of Interactive Mobile Technologies*, *17*(2), 42–61.

Farooq, M. S., Salam, M., Jaafar, N., Fayolle, A., Ayupp, K., Radovic-Markovic, M., & Sajid, A. (2017). Acceptance and use of lecture capture system (LCS) in executive business studies: Extending UTAUT2. *Interactive Technology and Smart Education*, *14*(4), 329–348.

Feng, Y. (Katherine), Claggett, J. L., Karahanna, E., & Tam, K. Y. (2022). A Randomized Field Experiment to Explore the Impact of Herding Cues as Catalysts for Adoption. *MIS Quarterly*, *46*(2), 1135–1163.

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business and Information Systems Engineering*, *66*(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Addison-Wesley.

Fishbein, M., & Ajzen, I. (2011). *Predicting and changing behavior: The reasoned action approach*. Psychology press.

Fjällräven. (2023). *Kånken Me*. https://www.fjallraven.com/us/en-us/bags-%0Agear/kanken/kanken-bags/kanken-me

Flamer, S. (1983). Assessment of the Multitrait-Multimethod Matrix Validity of Likert Scales via Confirmatory Factor Analysis. *Multivariate Behavioral Research*, *18*(3), 275–306.

Floridi, L., & Cowls, J. (2019). A Unifed Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1–15. https://doi.org/10.1162/99608f92.8cd550d1

Flynn, L. R., & Goldsmith, R. E. (1999). A Short, Reliable Measure of Subjective Knowledge. *Journal of Business Research*, *46*(1), 57–66. https://doi.org/10.1016/S0148-2963(98)00057-5

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models With Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39–50.

Frank, D. A., Chrysochou, P., & Mitkidis, P. (2023). The Paradox of Technology: Negativity Bias in Consumer Adoption of Innovative Technologies. *Psychology and Marketing*, *40*(3), 554–566. https://doi.org/10.1002/mar.21740

Franke, N., Schreier, M., & Kaiser, U. (2010). The “I Designed It Myself” Effect in Mass Customization. *Management Science*, *56*(1), 125–140. https://doi.org/10.1287/MNSC.1090.1077

Freudenstein, G., Zies, I., & Busche, S. (2019). *Der ewige Kampf der Banken mit der Legacy-IT*. Bain & Company Germany Inc. and Bain & Company Switzerland Inc.

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Will humans-in-the-loop become borgs? Merits and pitfalls of working with AI. *MIS Quarterly*, *45*(3), 1527–1556. https://doi.org/10.25300/MISQ/2021/16553

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2022). Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation. *Information Systems Research*, *33*(2), 678–696. https://doi.org/10.1287/isre.2021.1079

Fukuyama, F. (1995). *Trust: The Social Virtues and the Creation of Prosperity*. Free Press, New York.

Fuller, C. M., Simmering, M. J., Atinc, G., Atinc, Y., & Babin, B. J. (2016). Common Methods Variance Detection in Business Research. *Journal of Business Research*, *69*(8), 3192–3198.

Gartner. (2015). *Gartner’s 2015 Hype Cycle for Emerging Technologies Identifies the Computing Innovations That Organizations Should Monitor*. https://www.gartner.com/en/newsroom/press-releases/2015-08-18-gartners-2015-hype-cycle-for-emerging-technologies-identifies-the-computing-innovations-that-organizations-should-monitor

Gartner. (2022). *Was gibt es Neues im Hype Cycle 2022 von Gartner für neue Technologien?* https://www.gartner.de/de/artikel/was-ist-neu-im-hype-cycle-2022-von-gartner-fuer-neue-technologien

Gefen, D., Straub, D., & Boudreau, M.-C. (2000). Structural Equation Modeling and Regression:

Guidelines for Research Practice. *Communications of the Association for Information Systems*, *4*(October). https://doi.org/10.17705/1cais.00407

Gensler, S., Hinz, O., Skiera, B., & Theysohn, S. (2012). Willingness-To-Pay Estimation With Choice-Based Conjoint Analysis: Addressing Extreme Response Behavior With Individually Adapted Designs. *European Journal of Operational Research*, *219*(2), 368–378.

George, J. F., Gupta, M., Giordano, G., Mills, A. M., Tennant, V. M., & Lewis, C. C. (2018). The Effects of Communication Media and Culture on Deception Detection Accuracy. *MIS Quarterly*, *42*(2), 551–575.

Gezici, B., & Tarhan, A. K. (2022). Systematic Literature Review on Software Quality for AI-Based Software. *Empirical Software Engineering*, *27*(3), 1–65.

Gholami, F., Attar, N., Haghighi, H., Asl, M. V., Valueian, M., & Mohamadyari, S. (2018). A Classifier-Based Test Oracle for Embedded Software. *2018 Real-Time and Embedded Systems and Technologies (RTEST)*, 104–111.

Gibson, M., & Arnott, D. (2007). The Use of Focus Groups in Design Science Research. *ACIS 2007 Proceedings*, 327–337.

Gielnik, M. M., Spitzmuller, M., Schmitt, A., Klemann, D. K., & Frese, M. (2015). “I Put in Effort, Therefore I Am Passionate”: Investigating the Path From Effort to Passion in Entrepreneurship. *Academy of Management Journal*, *58*(4), 1012–1031.

Giessmann, A., & Stanoevska-Slabeva, K. (2013). What Are Developers’ Preferences on Platform as a Service? An Empirical Investigation. *2013 46th Hawaii International Conference on System Sciences*, 1035–1044.

Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT Outperforms Crowd-Workers for Text-Annotation Tasks. *ArXiv Preprint ArXiv:2303.15056*.

GitHub Blog. (2023). *Research: Quantifying GitHub Copilot’s impact on code quality*. https://github.blog/2023-10-10-research-quantifying-github-copilots-impact-on-code-quality/

GitHub Copilot. (2023). *Your AI pair programmer*. https://github.com/features/copilot

Goebel, P., Reuter, C., Pibernik, R., Sichtmann, C., & Bals, L. (2018). Purchasing Managers ‘ Willingness to Pay for Attributes That Constitute Sustainability. *Journal of Operations Management*, *62*, 44–58.

Goldman Sachs. (2023). *AI Investment Forecast to Approach $200 Billion Globally by 2025*. https://www.goldmansachs.com/insights/articles/ai-investment-forecast-to-approach-200-billion-globally-by-2025.html

Gondra, I. (2008). Applying Machine Learning to Software Fault-Proneness Prediction. *The Journal of Systems and Software*, *81*(2), 186–195.

Gong, X., Cheung, C. M. K., Zhang, K. Z. K., Chen, C., & Lee, M. K. O. (2021). A Dual-Identity Perspective of Obsessive Online Social Gaming. *Journal of the Association for Information Systems*, *22*(5), 1245–1284.

Goodman, S. A., & Svyantek, D. J. (1999). Person–Organization Fit and Contextual Performance: Do Shared Values Matter. *Journal of Vocational Behavior*, *55*(2), 254–275.

Göring, S., Rao, R. R. R., Merten, R., & Raake, A. (2023). Analysis of Appeal for Realistic AI-Generated Photos. *IEEE Access*, 38999–39012.

Green, P. E., & Rao, V. R. (1971). Conjoint Measurement for Quantifying Judgmental Data. *Journal of Marketing Research*, *8*(3), 355–363.

Green, P. E., & Srinivasan, V. (1978). Conjoint Analysis in Consumer Research Issues and Outlook. *Journal of Consumer Research*, *5*(2), 103–123.

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, *37*(2), 337–355.

Gu, M., Shu, C., & De Clercq, D. (2021). Knowledge Filters and Employee Venturing Behaviors: A Cross-Institutional Study of the U.S. and Indian Firms. *IEEE Transactions on Engineering Management*, 1–14.

Guan, C., Hung, Y. C., & Liu, W. (2022). Cultural Differences in Hospitality Service Evaluations: Mining Insights of User Generated Content. *Electronic Markets*, *32*, 1061–1081.

Gupta, M., Bastani, F., Khan, L., & Yen, I. L. (2004). Automated Test Data Generation Using MEA-Graph Planning. *16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2004)*, 174–182.

Guveyi, E., Aktas, M. S., & Kalipsiz, O. (2020). Human Factor on Software Quality: A Systematic Literature Review. *International Conference on Computational Science and Its Applications*, 918–930.

Haan, K. (2023). *Over 75% Of Consumers Are Concerned About Misinformation From Artificial Intelligence*. https://www.forbes.com/advisor/business/artificial-intelligence-consumer-sentiment/

Hair Jr., J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis* (6th ed.). Pearson Prentice Hall.

Hair Jr., J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (2nd ed.). SAGE Publications Inc.

Haleem, A., Javaid, M., & Singh, R. P. (2022). An Era of ChatGPT as a Significant Futuristic Support Tool: A Study on Features, Abilities, and Challenges. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, *2*(4), 100089.

Hall, E. T. (1976). *Beyond Culture*. Anchor Press, New York.

Hall, E. T., & Hall, M. R. (1990). *Understanding Cultural Differences*. Intercultural Press, Yarmouth, ME.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: Mturk Participants Perform Better on Online Attention Checks Than Do Subject Pool Participants. *Behavior Research Methods*, *48*(1), 400–407.

Henry, K. E., Kornfield, R., Sridharan, A., Linton, R. C., Groh, C., Wang, T., Wu, A., Mutlu, B., & Saria, S. (2022). Human–machine teaming is key to AI adoption: clinicians' experiences with a deployed machine learning system. *Npj Digital Medicine*, *5*(1), 1–6. https://doi.org/10.1038/s41746-022-00597-7

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A New Criterion for Assessing Discriminant Validityin Variance-Based Structural Equation Modeling. *Journal of the Academy of Marketing Science*, *43*, 115–135.

Hentzen, J. K., Hoffmann, A., Dolan, R., & Pala, E. (2021). Artificial Intelligence in Customer-Facing Financial Services: A Systematic Literature Review and Agenda for Future Research. *International Journal of Bank Marketing*, 1–38.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Research. *MIS Quarterly*, *28*(1), 75–105.

Hoehle, H., & Venkatesh, V. (2015). Mobile Application Usability: Conceptualization and Instrument Development. *MIS Quarterly*, *39*(2), 435–472.

Hoehle, H., Zhang, X., & Venkatesh, V. (2015). An Espoused Cultural Perspective to Understand Continued Intention to Use Mobile Applications: A Four-Country Study of Mobile Social Media Application Usability. *European Journal of Information Systems*, *24*, 337–359.

Hofstede, G. (1980). *Culture's Consequences: International Differences in Work-Related Values*. Sage, Beverly Hills, CA.

Hofstede, G. (1991). *Cultures and Organizations: Software of the Mind*. McGraw-Hill, London, UK.

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. Sage, Thousand Oaks, CA.

Hofstede, G. (2011). Dimensionalizing Cultures: The Hofstede Model in Context. *Online Readings in Psychology and Culture*, *2*.

Hofstede Insights. (2010). *Compare Countries*. https://www.hofstede-insights.com/fi/product/compare-countries/

Horani, O. M., Al-Adwan, A. S., Yaseen, H., Hmoud, H., Al-Rahmi, W. M., & Alkhalifah, A. (2023). The critical determinants impacting artificial intelligence adoption at the organizational level. *Information Development*. https://doi.org/10.1177/02666669231166889

Hourani, H., Hammad, A., & Lafi, M. (2019). The Impact of Artificial Intelligence on Software Testing. *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, 565–570.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*. Sage, Thousand Oaks, CA.

Hovav, A., Gnizy, I., & Han, J. (2021). The Effects of Cyber Regulations and Security Policies on Organizational Outcomes: A Knowledge Management Perspective. *European Journal of Information Systems*, 1–19.

Hristov, I., Camilli, R., & Mechelli, A. (2022). Cognitive Biases in Implementing a Performance Management System: Behavioral Strategy for Supporting Managers' Decision-Making Processes. *Management Research Review*, *45*(9), 1110–1136. https://doi.org/10.1108/MRR-11-2021-0777

Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *QUALITATIVE HEALTH RESEARCH*, *15*(9), 1277–1288.

Hunt, S. D., Sparkman, R. D., & Wilcox, J. B. (1982). The Pretest in Survey Research: Issues and Preliminary Findings. *Journal of Marketing Research*, *19*(2), 269.

IBM. (2022a). *Global Data from IBM Shows Steady AI Adoption as Organizations Look to Address Skills Shortages, Automate Processes and Encourage Sustainable Operations*. https://newsroom.ibm.com/2022-05-19-Global-Data-from-IBM-Shows-Steady-AI-Adoption-as-Organizations-Look-to-Address-Skills-Shortages,-Automate-Processes-and-Encourage-Sustainable-Operations

IBM. (2022b). *IBM Global AI Adoption Index 2022*. https://www.ibm.com/downloads/cas/GVAGA3JP

Inzlicht, M., & Campbell, A. V. (2022). Effort feels meaningful. *Trends in Cognitive Sciences*, *26*(12), 1035–1037.

Itkin, I., Novikov, A., & Yavorskiy, R. (2019). Development of Intelligent Virtual Assistant for Software Testing Team. *2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, 126–129.

Ives, B., & Jarvenpaa, S. L. (1991). Applications of Global Information Technology: Key Issues for Management. *MIS Quarterly*, *15*(1), 33–49.

Jachimowicz, J. M., & Weisman, H. (2022). Divergence Between Employer and Employee Understandings of Passion: Theory and Implications for Future Research. *Research in Organizational Behavior*, *42*(100167).

Jachimowicz, J. M., Wihler, A., Bailey, R. E., & Galinsky, A. D. (2018). Why Grit Requires

Perseverance and Passion to Positively Predict Performance. *Proceedings of the National Academy of Sciences*, *115*(40), 9980–9985.

Javidan, M., Dorfman, P., Sully de Luque, M., & House, R. J. (2006). In the Eye of the Beholder: Cross Cultural Lessons in Leadership From Project Globe. *The Academy of Management Perspective*, *20*, 67–90.

Jiang, Y., Li, X., Luo, H., Yin, S., & Kaynak, O. (2022). Quo vadis artificial intelligence? *Discover Artificial Intelligence*, *2*(1). https://doi.org/10.1007/s44163-022-00022-8

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or Not, AI Comes—an Interview Study of Organizational AI Readiness Factors. *Business & Information Systems Engineering*, *63*(1), 5–20.

Jordan, P. J., & Troth, A. C. (2020). Common Method Bias in Applied Settings: The Dilemma of Researching in Organizations. *Australian Journal of Management*, *45*(1), 3–14.

Joshi, A., Kale, S., Chandel, S., & Pal, D. (2015). Likert Scale: Explored and Explained. *British Journal of Applied Science & Technology*, *7*(4), 396–403.

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why Are We Averse Towards Algorithms ? A Comprehensive Literature Review on Algorithm Aversion. *ECIS 2020 Proceedings*, 1–16.

Jussupow, E., Spohrer, K., Heinzl, A., & Gawlitza, J. (2021). Augmenting Medical Diagnosis Decisions? An Investigation into Physicians' Decision-Making Process with Artificial Intelligence. *Information Systems Research*, *32*(3), 713–735. https://doi.org/10.1287/isre.2020.0980

Kahneman, D. (2012). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy*, *98*(6).

Kalish, S., & Nelson, P. (1991). A Comparison of Ranking, Rating and Reservation Price Measurement in Conjoint Analysis. *Marketing Letters*, *2*(4), 327–335.

Kamtarin, M. (2012). The Effect of Electronic Word of Mouth, Trust and Perceived Value on Behavioral Intention From the Perspective of Consumers. *International Journal of Academic Research in Economics and Management Sciences*, *1*(4), 56–66.

Kanngiesser, P., Gjersoe, N., & Hood, B. M. (2010). The effect of creative labor on property-ownership transfer by preschool children and adults. *Psychological Science*, *21*(9), 1236–1241. https://doi.org/10.1177/0956797610380701/ASSET/IMAGES/LARGE/10.1177_0956797610380701-FIG2.JPEG

Kanuri, V. K., Thorson, E., & Mantrala, M. K. (2014). Using Reader Preferences to Optimize News Content: A Method and a Case Study. *International Journal on Media Management*, *16*(2), 55–75.

Kapsis, I. (2020). Artificial Intelligence in Financial Services: Systemic Implications and Regulatory Responses. *Banking & Financial Services Policy Report*, 1–21.

Kar, S., Kar, A. K., & Gupta, M. P. (2021). Modeling Drivers and Barriers of Artificial Intelligence Adoption: Insights from a Strategic Management Perspective. *Intelligent Systems in Accounting, Finance & Management*, *28*(4), 217–238.

Kazmi, R., Jawawi, D. N. A., Mohamad, R., & Ghani, I. (2017). Effective Regression Test Case Selection: A Systematic Literature Review. *ACM Computing Surveys*, *50*(2), 1–32.

Ke, W., Kang, L., Tan, C. H., & Peng, C. H. (2021). User Competence With Enterprise Systems: The Effects of Work Environment Factors. *Information Systems Research*, *32*(3), 860–875.

Ke, W., & Zhang, P. (2011). Effects of Empowerment on Performance in Open-Source Software

Projects. *IEEE Transactions on Engineering Management*, *58*(2), 334–346. https://doi.org/10.1109/TEM.2010.2096510

Keil, M., Im, G. P., & Mähring, M. (2007). Reporting Bad News on Software Projects: The Effects of Culturally Constituted Views of Face-Saving. *Information Systems Journal*, *17*(1), 59–87.

Keil, M., Tan, B. C. Y., Wei, K.-K., Saarinen, T., Tuunainen, V., & Wassenaar, A. (2000). A Cross-Cultural Study on Escalation of Commitment Behavior in Software Projects. *MIS Quarterly*, *24*(2), 299–325.

Kesri, V., Nayak, A., & Ponnalagu, K. (2021). AutoKG - An Automotive Domain Knowledge Graph for Software Testing: A Position Paper. *2021 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*, 234–238.

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer Credit-Risk Models via Machine-Learning Algorithms. *Journal of Banking & Finance*, *34*(11), 2767–2787.

Khatibsyarbini, M., Isa, M. A., Jawawi, D. N. A., Shafie, M. L. M., Wan-Kadir, W. M. N., Hamed, H. N. A., & Suffian, M. D. M. (2021). Trend Application of Machine Learning in Test Case Prioritization: A Review on Techniques. *IEEE Access*, 166262–166282.

Kim, D. J. (2008). Self-Perception-Based Versus Transference-Based Trust Determinants in Computer-Mediated Transactions: A Cross-Cultural Comparison Study. *Journal of Management Information Systems*, *24*(4), 13–45.

Kim, D. J., Yim, M. S., Sugumaran, V., & Rao, H. R. (2016). Web Assurance Seal Services, Trust and Consumers' Concerns: An Investigation of E-commerce Transaction Intentions Across Two Nations. *European Journal of Information Systems*, *25*, 252–273.

Kim, H.-W., & Kankanhalli, A. (2009). Investigating User Resistance to Information Systems Implementation: A Status Quo Bias Perspective. *MIS Quarterly*, *33*(3), 567–582.

Kim, Y., Dykema, J., Stevenson, J., Black, P., & Moberg, D. P. (2019). Straightlining: Overview of Measurement, Comparison of Indicators, and Effects in Mail – Web Mixed-Mode Surveys. *Social Science Computer Review*, *37*(2), 214–233.

King, T. M., Arbon, J., Santiago, D., Adamo, D., Chin, W., & Shanmugam, R. (2019). AI for Testing Today and Tomorrow: Industry Perspectives. *2019 IEEE International Conference on Artificial Intelligence Testing (AITest)*, 81–88.

King, W. R. (2005). External Validity in IS Survey Research. *Communications of the Association for Information Systems*, *16*(1), 880–894.

Kinkel, S., Baumgartner, M., & Cherubini, E. (2021). Prerequisites for the Adoption of AI Technologies in Manufacturing: Evidence From a Worldwide Sample of Manufacturing Companies. *Technovation*, *110*, 1–17.

Kirk, R. E. (2009). Experimental Design. In *Sage handbook of quantitative methods in psychology* (pp. 23–45). https://doi.org/10.1016/B978-008044910-4.00431-4

Kisekka, V., & Goel, S. (2022). An Investigation of the Factors that Influence Job Performance During Extreme Events: The Role of Information Security Policies. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-022-10281-6

Kononenko, O., Baysal, O., & Godfrey, M. W. (2016). Code review quality: how developers see it. *Proceedings of the 38th International Conference on Software Engineering*, 1028–1038. https://doi.org/10.1145/2884781.2884840

Korteling, J. H., van de Boer-Visschedijk, G. C., Blankendaal, R. A. M., Boonekamp, R. C., & Eikelboom, A. R. (2021). Human- versus Artificial Intelligence. *Frontiers in Artificial Intelligence*, *4*, 1–13. https://doi.org/10.3389/frai.2021.622364

Korzynski, P., Mazurek, G., Altmann, A., Ejdys, J., Kazlauskaite, R., Paliszkiewicz, J., Wach, K., & Ziemba, E. (2023). Generative artificial intelligence as a new context for management theories: analysis of ChatGPT. *Central European Management Journal*, *31*(1), 3–13. https://doi.org/10.1108/CEMJ-02-2023-0091

Kranz, J. (2021). Strategic Innovation in IT Outsourcing: Exploring the Differential and Interaction Effects of Contractual and Relational Governance Mechanisms. *Journal of Strategic Information Systems*, *30*(1), 101656.

Krishnan, C., & Mariappan, J. (2024). The AI Revolution in E-Commerce: Personalization and Predictive Analytics. In L. Gaur & A. Abraham (Eds.), *Role of Explainable Artificial Intelligence in E-Commerce* (pp. 53–64). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-55615-9_4

Kruse, L., Wunderlich, N., & Beck, R. (2019). Artificial Intelligence for the Financial Services Industry: What Challenges Organizations to Succeed. *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Kuechler, W., & Vaishnavi, V. (2008). On Theory Development in Design Science Research: Anatomy of a Research Project. *European Journal of Information Systems*, *17*(5), 489–504.

Kumar, G., Kumar, R., Vedpal, S., Shivam, A., & Mishra, K. (2024). Analyzing Barriers in Adoption of Artificial Intelligence for Resilient Health Care Services to Society. *Global Journal of Flexible Systems Management*, *25*(1), 179–197. https://doi.org/10.1007/s40171-024-00373-4

Kung, L., Cegielski, C. G., & Kung, H. (2015). An Integrated Environmental Perspective on Software as a Service Adoption in Manufacturing and Retail Firm. *Journal of Information Technology*, 1–12.

Kushary, D., Davison, A. C., & Hinkley, D. V. (2000). Bootstrap Methods and Their Application. *Technometrics*, *42*(4), 216.

Lampo, A. (2022). How is Technology Accepted? Fundamental Works in User Technology Acceptance from Diffusion of Innovations to UTAUT-2. *ACM International Conference Proceeding Series*, 260–266. https://doi.org/10.1145/3568834.3568903

Larson, Z. (2023). *ChatGPT Isn't Coming for Your Coding Job*. https://www.wired.com/story/chatgpt-coding-software-crisis/

Le Gall-Ely, M. (2009). Definition, Measurement and Determinants of the Consumer's Willingness to Pay: A Critical Synthesis and Avenues for Further Research Marine. *Recherche et Applications En Marketing (English Edition)*, *24*(2), 91–112.

Lee, I., & Shin, Y. J. (2018). Fintech: Ecosystem, Business Models, Investment Decisions, and Challenges. *Business Horizons*, *61*, 35–46.

Lee, M. S., Geistfeld, L. V, & Stoel, L. (2007). Cultural Differences Between Korean and American Apparel Web Sites. *Journal of Fashion Marketing and Management*, *11*(4), 511–528.

Lee, Y. S., Kim, T., Choi, S., & Kim, W. (2022). When does AI pay off? AI -adoption intensity, complementary investments, and R&D strategy. *Technovation*, *118*(June), 102590. https://doi.org/10.1016/j.technovation.2022.102590

Leung, S.-O. (2011). A Comparison of Psychometric Properties and Normality in 4-, 5-, 6-, and 11-Point Likert Scales. *Journal of Social Service Research*, *37*(4), 412–421.

Li, X., & Prasad, C. (2005). Effectively Teaching Coding Standards in Programming. *Proceedings of the 6th Conference on {Information} Technology Education*, 239–244. https://doi.org/10.1145/1095714.1095770

Li, Z., Jing, X. Y., & Zhu, X. (2018). Progress on Approaches to Software Defect Prediction. *IET Software*, *12*(3), 161–175.

Liang, H., Wang, N., & Xue, Y. (2022). Juggling Information Technology (IT) Exploration and Exploitation: A Proportional Balance View of IT Ambidexterity. *Information Systems*

Research.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, *89*(1).

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*.

Lima, R., da Cruz, A. M. R., & Ribeiro, J. (2020). Artificial Intelligence Applied to Software Testing: A Literature Review. *2020 5th Iberian Conference on Information Systems and Technologies (CISTI)*.

Lin, J., Carter, L., Liu, D., Lin, J., Carter, L., & Liu, D. (2021). Privacy Concerns and Digital Government: Exploring Citizen Willingness to Adopt the Covidsafe App. *European Journal of Information Systems*, *30*(4), 389–402.

Ling, I. L., Liu, Y. F., Lin, C. W., & Shieh, C. H. (2020). Exploring IKEA Effect in Self-Expressive Mass Customization: Underlying Mechanism and Boundary Conditions. *Journal of Consumer Marketing*, *37*(4), 365–374. https://doi.org/10.1108/JCM-09-2017-2373/FULL/PDF

Liu, Liu, D., Santhanam, R., University of Oklahoma, Webster, J., & Queen's University. (2017). Toward Meaningful Engagement: A Framework for Design and Research of Gamified Information Systems. *MIS Quarterly*, *41*(4), 1011–1034. https://doi.org/10.25300/MISQ/2017/41.4.01

Liu, T., Vickers, B. D., & Seidler, R. D. (2023). Neural Correlates of Overvaluation and the Effort to Save Possessions in a Novel Decision Task: An Exploratory Fmri Study. *Frontiers in Psychology*, *14*, 1059051. https://doi.org/10.3389/FPSYG.2023.1059051/BIBTEX

Liu, Y., Le-Cong, T., Widyasari, R., Tantithamthavorn, C., Li, L., Le, X.-B. D., & Lo, D. (2023). *Refining ChatGPT-Generated Code: Characterizing and Mitigating Code Quality Issues*. arXiv. http://arxiv.org/abs/2307.12596

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality : Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, *74*(5), 570–581. https://doi.org/10.1002/asi.24750

Lund Research Ltd. (2012). *Constructs in Quantitative Research*. https://dissertation.laerd.com/constructs-in-quantitative-research.php#:~:text=Instead%2C we use variables to operationalize %28i.e.%2C measure%29,the same %28e.g.%2C the construct and variable%2C sex%29.

Lysova, E. I., Tosti-Kharas, J., Michaelson, C., Fletcher, L., Bailey, C., & McGhee, P. (2023). Ethics and the Future of Meaningful Work: Introduction to the Special Issue. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-023-05345-9

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques. *MIS Quarterly*, *35*(2), 293–334.

Madera, M., & Tomoń, R. (2017). A Case Study on Machine Learning Model for Code Review Expert System in Software Engineering. *Federated Conference on Computer Science and Information Systems (FedCSIS)*, 1357–1363.

Maier, C., Laumer, S., Tarafdar, M., Mattke, J., Reis, L., & Weitzel, T. (2021). Challenge and Hindrance IS Use Stressors and Appraisals: Explaining Contrarian Associations in Post-Acceptance IS Use Behavior. *Journal of the Association for Information Systems*, *22*(6), 1590–1624.

Maier, C., Thatcher, J. B., Grover, V., & Dwivedi, Y. K. (2023). Cross-Sectional Research: A Critical Perspective, Use Cases, and Recommendations for Is Research. *International Journal of Information Management*, *70*.

Makki, A., & Abid, M. (2017). Influence of Intrinsic and Extrinsic Motivation on Employee's Task Performance. *Studies in Asian Social Science*, *4*(1), 38. https://doi.org/10.5430/sass.v4n1p38

Marselis, R. (2020). Testing in the Digital Age. In S. Goericke (Ed.), *The Future of Software Quality Assurance* (pp. 77–92). Springer, Cham.

Marsh, L. E., Gil, J., & Kanngiesser, P. (2022). The Influence of Collaboration and Culture on the IKEA Effect: Does Cocreation Alter Perceptions of Value in British and Indian Children? *Developmental Psychology* , *58*(4), 662. https://doi.org/10.1037/dev0001321.supp

Marsh, L. E., Kanngiesser, P., & Hood, B. (2018). When and How Does Labour Lead to Love? The Ontogeny and Mechanisms of the IKEA Effect. *Cognition*, *170*, 245–253. https://doi.org/10.1016/J.COGNITION.2017.10.012

Matell, M. S., & Jacoby, J. (1971). Is There an Optimal Number of Alternatives for Likert Scale Items? Study I: Reliability and Validity. *Educational and Psychological Measurement*, *31*, 657–674.

May, D. R., Gilson, R. L., & Harter, L. M. (2004). The Psychological Conditions of Meaningfulness, Safety and Availability and the Engagement of the Human Spirit at Work. *Journal of Occupational and Organizational Psychology*, *77*(1), 11–37. https://doi.org/10.1348/096317904322915892

Mayring, P. (2000). Qualitative Content Analysis. *Forum: Qualitative Social Research*, *1*(2).

Mazaheri, E., Lagzian, M., & Hemmat, Z. (2020). Research Directions in Information Systems Field, Current Status and Future Trends: A Literature Analysis of AIS Basket of Top Journals. *Australasian Journal of Information Systems*, *24*, 1–25.

McCoy, S., Galletta, D. F., & King, W. R. (2007). Applying TAM Across Cultures: The Need for Caution. *European Journal of Information Systems*, *16*(1), 81–90.

McElheran, K., Li, J. F., Brynjolfsson, E., Kroff, Z., Dinlersoz, E., Foster, L., & Zolas, N. (2024). AI adoption in America: Who, what, and where. *Journal of Economics and Management Strategy*, *33*(2), 375–415. https://doi.org/10.1111/jems.12576

McKinsey. (2022). *Mind the Gap*. https://www.mckinsey.com/~/media/mckinsey/email/genz/2023/04/2023-04-25b.html

McKinsey. (2023). *The Economic Potential of Generative AI: The Next Productivity Frontier*. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#key-insights

Melnyk, S. A., Page, T. J., Wu, S. J., & Burns, L. A. (2012). Would You Mind Completing This Survey: Assessing the State of Survey Research in Supply Chain Management. *Journal of Purchasing and Supply Management*, *18*(1), 35–45.

Memon, A. M., Pollack, M. E., & Soffa, M. Lou. (2001). Hierarchical Gui Test Case Generation Using Automated Planning. *IEEE Transactions on Software Engineering*, *27*(2), 144–155.

Mihale-Wilson, A. C., Zibuschka, J., & Hinz, O. (2019). User Preferences and Willingness to Pay For In-Vehicle Assistance. *Electronic Markets*, *29*(1), 37–53.

Miller, D. T., & Ross, M. (1975). Self-Serving Biases in the Attribution of Causality: Fact or Fiction? *Psychological Bulletin*, *82*(2), 213–225. https://doi.org/10.1037/h0076486

Miltgen, C. L., & Peyrat-Guillard, D. (2014). Cultural and Generational Influences on Privacy Concerns: A Qualitative Study in Seven European Countries. *European Journal of Information Systems*, *23*(2), 103–125.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

Mochon, D., Norton, M. I., & Ariely, D. (2012). Bolstering and Restoring Feelings of Competence via the IKEA Effect. *International Journal of Research in Marketing*, *29*(4), 363–369. https://doi.org/10.1016/J.IJRESMAR.2012.05.001

Mohr, S., & Kühl, R. (2021). Acceptance of artificial intelligence in German agriculture : an application of the technology acceptance model and the theory of planned behavior. *Precision Agriculture*, *22*(6), 1816–1844. https://doi.org/10.1007/s11119-021-09814-x

Momani, A. M., & Jamous, M. M. (2017). *The Evolution of Technology Acceptance Theories*. *1*(1), 51–58.

Moore, G. C., & Benbasat, I. (1991). Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. *Information Systems Research*, *2*(3), 192–222.

Moorman, C. (2008). Cross-Sectional Versus Longitudinal Survey Research: Concepts, Findings, and Guidelines. *Journal of Marketing Research*, *45*, 261–279.

Moorthy, S., Ratchford, B. T., & Talukdar, D. (1997). Consumer Information Search Revisited: Theory and Empirical Analysis. *Journal of Consumer Research*, *23*(4), 263–277.

Mortimer, S. A. (2023). What Makes Work Meaningful? *Journal of Business Ethics*, *185*, 835–845.

Moser, C., & Deichmann, D. (2021). Knowledge Sharing in Two Cultures: The Moderating Effect of National Culture on Perceived Knowledge Quality in Online Communities. *European Journal of Information Systems*, *30*(6), 623–641.

Motowidlo, S. J., & Scotter, J. R. Van. (1994). Evidence That Task Performance Should Be Distinguished From Contextual Performance. *Journal of Applied Psychology*, *79*(4), 475.

Myers, M. D., & Newman, M. (2007). The Qualitative Interview in IS Research: Examining the Craft. *Information and Organization*, *17*(1), 2–26.

Myers, M. D., & Tan, F. B. (2002). Beyond Models of National Culture in Information Systems Research. *Journal of Global Information Management*, *10*(1), 24–32.

Newsted, P. R., Huff, S. L., & Munro, M. C. (1998). Survey Instruments in Information Systems. *MIS Quarterly*, *22*(4), 553–554.

Nguyen, K. A. T., Nguyen, T. A. T., Nguelifack, B. M., & Jolly, C. M. (2022). Machine Learning Approaches for Predicting Willingness to Pay For Shrimp Insurance in Vietnam. *Marine Resources Economics*, *37*(2), 155–182.

Ni, F., Arnott, D., & Gao, S. (2019). The Anchoring Effect in Business Intelligence Supported Decision-Making. *Journal of Decision Systems*, *28*(2), 67–81. https://doi.org/10.1080/12460125.2019.1620573

Nickerson, R. C., Muntermann, J., Varshney, U., & Isaac, H. (2009). *Taxonomy Development In Information Systems: Developing A Taxonomy Of Mobile Applications*. https://www.researchgate.net/publication/46479041

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. https://doi.org/10.1057/ejis.2012.26

Northwest Executive Education. (2023). *Top 8 ChatGPT Use Cases for Businesses*. https://northwest.education/insights/careers/top-8-chatgpt-use-cases-for-businesses/

Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA Effect: When Labor Leads to Love. *Journal of Consumer Psychology*, *22*(3), 453–460. https://doi.org/10.1016/J.JCPS.2011.08.002

Novikov, A. M., & Novikov, D. A. (2013). Research methodology: From philosophy of science to research design. In *Kybernetes* (Vol. 41, Issue 5/6). Crc Press. https://doi.org/10.1108/k.2012.06741eaa.005

Noy, S., & Zhang, W. (2023). Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4375283

Nunamaker Jr., J. F., Briggs, R. O., Derrick, D. C., & Schwabe, G. (2015). The Last Research Mile: Achieving Both Rigor and Relevance in Information Systems Research. *Journal of Management Information Systems*, *32*(3), 10–47. https://doi.org/10.1080/07421222.2015.1094961

Nuñez-Varela, A. S., Pérez-Gonzalez, H. G., Martínez-Perez, F. E., & Soubervielle-Montalvo, C. (2017). Source Code Metrics: A Systematic Mapping Study. *The Journal of Systems and Software*, *128*, 164–197.

O'Nyumba, T., Wilson, K., Derrick, C. J., & Mukherjee, N. (2018). The Use of Focus Group Discussion Methodology: Insights From Two Decades of Application in Conservation. *Methods in Ecology and Evolution*, *9*(1), 20–32.

Odean, T. (1998). Volume, Volatility, Price, and Profit When All Traders Are Above Average. *Journal of Finance*, *53*(6), 1887–1934. https://doi.org/10.1111/0022-1082.00078

OECD. (2020). *Trade Interdependencies in Covid-19 Goods*. https://www.oecd.org/coronavirus/policy-responses/trade-interdependencies-in-covid-19-goods-79aaa1d6/

Ojer, M., Serrano, I., Saiz, F., Barandiaran, I., Gil, I., Aguinaga, D., & Alejandro, D. (2020). Real-Time Automatic Optical System to Assist Operators in the Assembling of Electronic Components. *The International Journal of Advanced Manufacturing Technology*, *107*(5), 2261–2275.

Okoli, C., & Schabram, K. (2010). A Guide to Conducting a Systematic Literature Review of Information Systems Research. *Sprouts: Working Papers on Information Systems*, *10*(26).

Oliveira, T., & Martins, M. F. (2011). Literature Review of Information Technology Adoption Models at Firm Level. *Electronic Journal of Information Systems Evaluation*, *14*(1), 110–121.

OpenAI. (2022). *Introducing ChatGPT*. https://openai.com/index/chatgpt/

OpenAI. (2023a). *Dall-E 3*. https://openai.com/dall-e-3

OpenAI. (2023b). *Language Models can Explain Neurons in Language Models*. https://openai.com/research/language-models-can-explain-neurons-in-language-models

Paas, L. J., & Morren, M. (2018). Please Do Not Answer if You Are Reading This: Respondent Attention in Online Panels. *Marketing Letters*, *29*(1), 13–21.

Pan, R., Bagherzadeh, M., Ghaleb, T. A., & Briand, L. (2022). Test Case Selection and Prioritization Using Machine Learning: A Systematic Literature Review. *Empirical Software Engineering*, *27*(2), 1–43.

Pandey, S. K., Rathee, D., & Tripathi, A. K. (2020). Software Defect Prediction Using K-PCA and Various Kernel-Based Extreme Learning Machine: An Empirical Study. *IET Software*, *14*(7), 768–782.

Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77.

Peres, R. S., Jia, X., Lee, J., Sun, K., Colombo, A. W., & Barata, J. (2020). Industrial Artificial Intelligence in Industry 4.0-Systematic Review, Challenges and Outlook. *IEEE Access*, *8*, 220121–220139.

Perfecto. (2023). *Perfecto Home*. https://www.perfecto.io/

Perkel, J. M. (2023). Six tips for better coding with {ChatGPT}. *Nature*, *618*(7964), 422–423. https://doi.org/10.1038/d41586-023-01833-0

Peter, J. P. (1981). Construct Validity: A Review of Basic Issues and Marketing Practices. *Journal* *of Marketing Research*, *18*(2), 133.

Peters, F., Pumplun, L., & Buxmann, P. (2020). Opening the Black Box: Consumer's Willingness to Pay for Transparency of Intelligent Systems. *Proceedings of the 28th European Conference on Information Systems*, 90. https://aisel.aisnet.org/ecis2020_rp/90

Petter, S., Straub, D., & Rai, A. (2007). Specifying Formative Constructs in Information Systems Research. *MIS Quarterly*, *31*(4), 623–656.

Phillips-Wren, G., Jefferson, T., & McKniff, S. (2019). Cognitive Bias and Decision Aid Use Under Stressful Conditions. *Journal of Decision Systems*, *28*(2), 162–184. https://doi.org/10.1080/12460125.2019.1643695

Piehlmaier, D. M. (2022). Overconfidence and the Adoption of Robo-Advice: Why Overconfident Investors Drive the Expansion of Automated Financial Advice. *Financial Innovation*. https://doi.org/10.1186/s40854-021-00324-3

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*, *88*(5), 879–903.

Polisetty, A., Chakraborty, D., Sowmya, G., & Kar, A. K. (2024). What Determines AI Adoption in Companies ? Mixed-Method Evidence What Determines AI Adoption in Companies ? Mixed-Method Evidence. *Journal of Computer Information Systems*, *64*(3), 370–387. https://doi.org/10.1080/08874417.2023.2219668

Polites, G. L., Roberts, N., & Thatcher, J. (2012). Conceptualizing Models Using Multidimensional Constructs: A Review and Guidelines for Their Use. *European Journal of Information Systems*, *21*(1), 22–48.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. Free Press.

Poth, A., Beck, Q., & Riel, A. (2019). Artificial Intelligence Helps Making Quality Assurance Processes Leaner. *European Conference on Software Process Improvement*, 722–730.

Prasad, S. G., Sharmila, V. C., & Badrinarayanan, M. K. (2023). Role of Artificial Intelligence based Chat Generative Pre-trained Transformer (ChatGPT) in Cyber Security. *Proceedings of the Second International Conference on Applied Artificial Intelligence and Computing (ICAAIC 2023)*, 107–114. https://doi.org/10.1109/ICAAIC56838.2023.10141395

Pumplun, L., Tauchert, C., & Heidt, M. (2019). A New Organizational Chassis for Artificial Intelligence-Exploring Organizational Readiness Factors. *Proceedings of the 27th European Conference on Information Systems (ECIS)*, 1–15.

Pumplun, Luisa, Fecho, M., Wahl, N., Peters, F., & Buxmann, P. (2021). Adoption of Machine Learning Systems for Medical Diagnostics in Clinics: Qualitative Interview Study. *Journal of Medical Internet Research*, *23*(10), e29301. https://doi.org/10.2196/29301 PM - 34652275

PwC. (2017). *Sizing the Prize. What's the Real Value of AI for Your Business and How Can You Capitalise?* https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf,

Rabbi, K., Mamun, Q., & Islam, M. R. (2018). An Efficient Particle Swarm Intelligence Based Strategy to Generate Optimum Test Data in T-Way Testing. *2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA)*, 123–128.

Radhakrishnan, J., & Chattopadhyay, M. (2020). Determinants and Barriers of Artificial Intelligence Adoption – A Literature Review. *IFIP Advances in Information and Communication Technology*, *617*, 89–99. https://doi.org/10.1007/978-3-030-64849-7_9

Radtke, T., Liszewska, N., Horodyska, K., Boberska, M., Schenkel, K., & Luszczynska, A. (2019).

Cooking Together: The IKEA Effect on Family Vegetable Intake. *British Journal of Health Psychology*, *24*(4), 896–912. https://doi.org/10.1111/BJHP.12385

Raghoebar, S., van Kleef, E., & de Vet, E. (2017). Self-Crafting Vegetable Snacks: Testing the Ikea-Effect in Children. *British Food Journal*, *119*(6), 1301–1312. https://doi.org/10.1108/BFJ-09-2016-0443/FULL/PDF

Rai, A., Tang, X., Yin, Z., & Du, S. (2022). Gaining Customer Loyalty with Tracking Information Quality in B2B Logistics. *Journal of Management Information Systems*, *39*(2), 307–335.

Ramos-Villagrasa, P. J., Barrada, J. R., Fernández-Del-Río, E., & Koopmans, L. (2019). Assessing Job Performance Using Brief Self-report Scales: The Case of the Individual Work Performance Questionnaire. *Journal of Work and Organizational Psychology*, *35*(3), 195–205.

Ramsey, S. M., & Bergtold, J. S. (2021). Examining Inferences From Neural Network Estimators of Binary Choice Processes: Marginal Effects, and Willingness-To-Pay. *Computational Economics*, *58*(4), 1137–1165.

Randall, T., Terwiesch, C., & Ulrich, K. T. (2007). Research Note—User Design of Customized Products. *Marketing Science*, *26*(2), 268–280.

Rao, A., Kim, J., Kamineni, M., Pag, M., Lie, W., & Succi, M. D. (2023). Evaluating ChatGPT as an Adjunct for Radiologic Decision-Making. *MedRxiv*.

Rathore, S. S., & Kumar, S. (2021). Software Fault Prediction Based on the Dynamic Selection of Learning Technique: Findings From the Eclipse Project Study. *Applied Intelligence*, *51*(12), 8945–8960.

Recker, J. (2021). *Scientific Research in Information Systems: A Beginner's Guide* (2nd ed.). Springer Nature Switzerland AG.

Riemenschneider, C. K., & Armstrong, D. J. (2021). The Development of the Perceived Distinctiveness Antecedent Of Information Systems Professional Identity. *MIS Quarterly*, *45*(3), 1149–1186.

Riikkinen, M., Saarijärvi, H., Sarlin, P., & Lähteenmäki, I. (2018). Using Artificial Intelligence to Create Value in Insurance. *International Journal of Bank Marketing*, 1145–1168.

Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). Free Press.

Roh, T., Park, B. Il, & Xiao, S. (Simon). (2023). Adoption of Ai-Enabled Robo-Advisors in Fintech: Simultaneous Employment of Utaut and the Theory of Reasoned Action. *Journal of Electronic Commerce Research*, *24*(1), 29–47.

Roßnagel, H., Zibuschka, J., Hinz, O., & Muntermann, J. (2014). Users' Willingness to Pay for Web Identity Management Systems. *European Journal of Information Systems*, *23*(1), 36–50.

Russel, S. J., & Norwig, P. (2021). *Artificial Intelligence: A Modern Approach*. Pearson Education.

Russo, D. (2024). Navigating the Complexity of Generative AI Adoption in Software Engineering. *ACM Transactions on Software Engineering and Methodology*, *33*(5). https://doi.org/10.1145/3652154

Samek, W., Wiegand, T., & Müller, K.-R. (2017). Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. *ITU Journal: ICT Discoveries*, 1–8.

Sande, J. B., & Ghosh, M. (2018). Endogeneity in Survey Research. *International Journal of Research in Marketing*, *35*(2), 185–204.

Sauvola, J., Tarkoma, S., Klemettinen, M., Riekki, J., & Doermann, D. (2024). Future of software development with generative AI. *Automated Software Engineering*, *31*(1), 1–8. https://doi.org/10.1007/s10515-024-00426-z

Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards Artificial

Intelligence Scale. *Computers in Human Behavior Reports*, *1*(February), 100014. https://doi.org/10.1016/j.chbr.2020.100014

Schewe, C. D., & Dillon, W. R. (1978). Marketing information system utilization: an application of self-concept theory. *Journal of Business Research*, *6*(1), 67–79.

Schieferdecker, I. (2020). Responsible Software Engineering. In *The Future of Software Quality Assurance* (pp. 137–146). Springer, Cham.

Schmidt, R., Lyytinen, K., Keil, M., & Cule, P. (2001). Identifying Software Project Risks: An International Delphi Study. *Journal of Management Information Systems*, *17*(4), 5–36.

Schmitz, K., Storey, V. C., Schmitz, K., & Robinson, J. M. (2020). Empirical Test Guidelines for Content Validity: Wash, Rinse, and Repeat until Clean. *Communications of the Association for Information Systems*, *47*, 787–850.

Schryen, G. (2015). Writing Qualitative IS Literature Reviews—Guidelines for Synthesis, Interpretation, and Guidance of Research. *Communications of the Association for Information Systems*, *37*(12), 286–325.

Schuir, J., & Teuteberg, F. (2021). Understanding Augmented Reality Adoption Trade-Offs in Production Environments From the Perspective of Future Employees: A Choice-Based Conjoint Study. *Information Systems and E-Business Management*, *19*(3), 1039–1085.

Segura, S., & Zhou, Z. Q. (2018). Metamorphic Testing 20 Years Later: A Hands-on Introduction. *2018 ACM/IEEE 40th International Conference on Software Engineering: Companion Proceeedings*, 538–539.

Serenko, A., & Turel, O. (2021). Why Are Women Underrepresented in the American IT Industry? The Role of Explicit and Implicit Gender Identities. *Journal of the Association for Information Systems*, *22*(1), 41–66.

Seymour, M., Yuan, L. I., Dennis, A. R., & Riemer, K. (2021). Have We Crossed the Uncanny Valley? Understanding Affinity, Trustworthiness, and Preference for Realistic Digital Humans in Immersive Environments. *Journal of the Association for Information Systems*, *22*(3), 591–617.

Shafiq, S., Mashkoor, A., Mayr-Dorn, C., & Egyed, A. (2021). A Literature Review of Using Machine Learning in Software Development Life Cycle Stages. *IEEE Access*, *9*, 140896–140920.

Shapiro, L. (2004). *Something from the oven: Reinventing dinner in 1950s America*. New York: Viking.

Sharma, P. N., Daniel, S. L., Chung, T. R., & Grover, V. (2022). A Motivation-Hygiene Model of Open Source Software Code Contribution and Growth. *Journal of the Association for Information Systems*, *23*(1), 165–195.

Sharma, S., Islam, N., Singh, G., & Dhir, A. (2022). Why Do Retail Customers Adopt Artificial Intelligence (AI) Based Autonomous Decision-Making Systems? *IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT*, 1–16.

Shen, P., Ding, X., Ren, W., & Yang, C. (2018). Research on Software Quality Assurance Based on Software Quality Standards and Technology Management. *2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 385–390.

Shinde, P. P., & Shah, S. (2018). A Review of Machine Learning and Deep Learning Applications. *Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018*, 1–6. https://doi.org/10.1109/ICCUBEA.2018.8697857

Shore, B., & Venkatachalam, A. R. (1996). Role of National Culture in the Transfer of Information Technology. *The Journal of Strategic Information Systems*, *5*(1), 19–35.

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational Decision-Making

Structures in the Age of Artificial Intelligence. *California Management Review*, *61*(4), 66–83. https://doi.org/10.1177/0008125619862257

Sia, C. L., Lim, K. H., Leung, K., Lee, M. K. O., Huang, W. W., & Benbasat, I. (2009). Web Strategies to Promote Internet Shopping: Is Cultural-Customization Needed? *MIS Quarterly*, *33*(3), 491–512.

Sinn, W., Thoben, S., & Graf, C. (2021). *Deutschlands Banken 2021: Wie die Renditewende gelingt*. Bain & Company Germany Inc. and Bain & Company Switzerland Inc.

Sjödin, D., Parida, V., Palmié, M., & Wincent, J. (2021). How AI Capabilities Enable Business Model Innovation: Scaling AI Through Co-evolutionary Processes and Feedback Loops. *Journal of Business Research*, *134*, 574–587.

Smith, E. R., & Conrey, F. R. (2007). Agent-Based Modeling: A New Approach for Theory Building in Social Psychology. *Personality and Social Psychology Review*, *11*(1), 87–104. https://doi.org/10.1177/1088868306294789

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information Privacy: Measuring Individuals' Concerns About Organizational Practices. *MIS Quarterly*, *20*(2), 167–195.

Sofy. (2023). *Pricing*. https://sofy.ai/pricing/

Sohn, K., & Kwon, O. (2020). Telematics and Informatics Technology acceptance theories and factors in fl uencing arti fi cial Intelligence-based intelligent products. *Telematics and Informatics*, *47*(April 2019), 1–14. https://doi.org/10.1016/j.tele.2019.101324

Sohn, K., Sung, C. E., Koo, G., & Kwon, O. (2021). Artificial Intelligence in the Fashion Industry: Consumer Responses to Generative Adversarial Network (GAN) Technology. *International Journal of Retail and Distribution Management*, *49*(1), 61–80.

Srite, M., & Karahanna, E. (2006). The Role of Espoused National Cultural Values in Technology Acceptance. *MIS Quarterly*, *30*(3), 679–704.

Srivastava, S. C., Chandra, S., & Shirish, A. (2015). Technostress Creators and Job Outcomes: Theorising the Moderating Influence Of Personality Traits. *Information Systems Journal*, *25*(4), 355–401.

Stahl, A. (2021). *How AI Will Impact The Future Of Work And Life*. https://www.forbes.com/sites/ashleystahl/2021/03/10/how-ai-will-impact-the-future-of-work-and-life/

Stahl, B. C., & Eke, D. (2024). The Ethics of ChatGPT – Exploring the Ethical Issues of an Emerging Technology. *International Journal of Information Management*, *74*. https://doi.org/10.1016/j.ijinfomgt.2023.102700

Statista. (2024). *Artificial Intelligence*. https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide

Steelman, Z. R., & Hammer, B. I. (2014). Data Collection in the Digital Age: Innovative Alternatives to Student Samples. *MIS Quarterly*, *38*(2), 355–378.

Stegeman, M., Barendsen, E., & Smetsers, S. (2014). Towards an Empirically Validated Model for Assessment of Code Quality. *Proceedings of the 14th Koli Calling International Conference on Computing Education Research*, 99–108. https://doi.org/10.1145/2674683.2674702

Stone, E. (2010). T test, paired samples. *Encyclopedia of Research Design*.

Storey, M.-A., Houck, B., & Zimmermann, T. (2022). How Developers and Managers Define and Trade Productivity for Quality. *15th International Conference on Cooperative and Human Aspects of Software Engineering (CHASE'22), May 21â•fi29, 2022, Pittsburgh, PA, USA*, 26–35.

Straub, D., Gefen, D., & Recker, J. (2022). *Quantitative Research in Information Systems*.

Association for Information Systems (AISWorld) Section on IS Research, Methods, and Theories, Last Updated March 25, 2022. http://www.janrecker.com/quantitative-research-in-information-systems/

Sturm, T., Gerlach, J. P., Pumplun, L., Mesbah, N., Peters, F., Tauchert, C., Nan, N., & Buxmann, P. (2021). Coordinating human and machine learning for effective organizational learning. *MIS Quarterly*, *45*(3), 1581–1602.

Sturm, T., & Peters, F. (2020). The Impact of Artificial Intelligence on Individual Performance: Exploring the Fit between Task , Data , and Technology THE IMPACT OF ARTIFICIAL INTELLIGENCE ON INDIVIDUAL PERFORMANCE : EXPLORING THE FIT BE- TWEEN TASK , DATA , AND TECHNOLOGY. *ECIS 2020 Proceedings*.

Subhash, S., Srivatsa, P. N., Siddesh, S., Ullas, A., & Santhosh, B. (2020). Artificial Intelligence-based Voice Assistant. *Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, 593–596.

Surameery, N. M. S., & Shakor, M. Y. (2023). Use ChatGPT to Solve Programming Bugs. *International Journal of Information Technology and Computer Engineering*, *31*, 17–22. https://doi.org/10.55529/ijitc.31.17.22

Tahvili, S., Afzal, W., Saadatmand, M., Bohlin, M., & Ameerjan, S. H. (2018). ESPRET: Espret: A Tool for Execution Time Estimation of Manual Test Cases. *The Journal of Systems & Software*, *146*, 26–41.

Teubner, T., Flath, C. M., Weinhardt, C., van der Aalst, W., & Hinz, O. (2023). Welcome to the era of chatgpt et al. the prospects of large language models. *Business & Information Systems Engineering*, *65*(2), 95–101.

Tremblay, M. C., Hevner, A. R., & Berndt, D. J. (2010). Focus Groups for Artifact Refinement and Evaluation in Design Research. *Communications of the Association for Information Systems*, *26*.

Trinkle, B. S., Warkentin, M., Malimage, K., & Raddatz, N. (2021). High-Risk Deviant Decisions: Does Neutralization Still Play a Role? *Journal of the Association for Information Systems*, *22*(3), 797–826.

Trompenaars, F., & Hampden-Turner, C. (1997). *Riding the Waves of Culture: Understanding Cultural Diversity in Business*. Nicholas Brealey Pub.

Trompenaars, F., & Woolliams, P. (2003). A New Framework for Managing Change. *Journal of Change Management*, *3*(4), 361–375.

Turel, O., Serenko, A., & Bontis, N. (2007). User Acceptance of Wireless Short Messaging Services: Deconstructing Perceived Value. *Information & Management*, *44*(1), 63–73. https://doi.org/10.1016/J.IM.2006.10.005

Turel, O., Yuan, Y., & Connelly, C. E. (2008). In Justice We Trust: Predicting User Acceptance of E-Customer Services. *Journal of Management Information Systems*, *24*(4), 123–151. https://doi.org/10.2753/MIS0742-1222240405

Tuunanen, T., & Kuo, I. Te. (2015). The Effect of Culture on Requirements: A Value-Based View of Prioritization. *European Journal of Information Systems*, *24*, 295–313.

Tversky, A., & Kahneman, D. (1973). Availability: A Heuristic for Judging Frequency. *COGNITTIVE PSYCHOLOG*, *5*, 207–232.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty : Heuristics and Biases. *Science*, *185*, 1124–1131.

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, *211*, 453–458.

United States Department of Labor. (n.d.). *SIC Manual*. Retrieved July 29, 2022, from https://www.osha.gov/data/sic-manual

Urquhart, C., Lehmann, H., & Myers, M. D. (2010). Putting the 'Theory' Back Into Grounded Theory: Guidelines for Grounded Theory Studies in Information Systems. *Information Systems Journal*, *20*(4), 357–381.

Uzumcu, O., & Acilmis, H. (2024). Do Innovative Teachers use AI-powered Tools More Interactively? A Study in the Context of Diffusion of Innovation Theory. *Technology, Knowledge and Learning*, *29*(2), 1109–1128. https://doi.org/10.1007/s10758-023-09687-1

Vance, A., Elie-dit-cosaque, C., & Straub, D. W. (2008). Examining Trust in Information Technology Artifacts: The Effects of System Quality and Culture. *Journal of Management Information Systems*, *24*(4), 73–100.

Venable, J., & Baskerville, R. (2012). Eating Our Own Cooking: Toward a More Rigorous Design Science of Research Methods. *Electronic Journal of Business Research Methods*, *10*(2), 141–153.

Venaik, S., & Brewer, P. (2013). Critical Issues in the Hofstede and GLOBE National Culture Models. *International Marketing Review*, *30*(5), 469–482.

Venkatesh, V. (2022). Adoption and Use of AI Tools: A Research Agenda Grounded in Utaut . *Annals of Operations Research*, *308*, 641–652. https://doi.org/10.1007/s10479-020-03918-9

Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, *39*(2), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the Qualitative–Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. *MIS Quarterly*, *37*(1), 21–54.

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model : Four Longitudinal Field Studies. *Management Science*, *46*(2), 186–204.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425–478.

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*, *36*(1), 157–178.

VERBI Software. (2022). *MAXQDA 2022*. www.maxqda.com

VHB. (2024). *VHB Publication Media Rating 2024*. https://vhbonline.org/fileadmin/user_upload/VHB_Rating_2024_Area_rating_WI.pdf

vom Brocke, J., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature. *ECIS 2009 Proceedings*.

von Wedel, P., & Hagist, C. (2022). Physicians' Preferences and Willingness to Pay For Artificial Intelligence-Based Assistance Tools: A Discrete Choice Experiment Among German Radiologists. *BMC Health Services Research*, *22*(1), 1–14.

Voorhees, C. M., Brady, M. K., Calantone, R., Ramirez, E., & Brady, M. K. (2016). Discriminant Validity Testing in Marketing: An Analysis, Causes for Concern, and Proposed Remedies. *Journal of the Academy of Markting Science*, *44*, 119–134.

VOSViewer. (2023). *Welcome to VOSViewer*. https://www.vosviewer.com/

Walasek, L., Rakow, T., & Matthews, W. J. (2017). When Does Construction Enhance Product

Value? Investigating the Combined Effects of Object Assembly and Ownership on Valuation. *Journal of Behavioral Decision Making*, *30*(2), 144–156. https://doi.org/10.1002/BDM.1931

Walgude, A., & Natarajan, S. (2019). *World Quality Report*. Capgemini. https://www.capgemini.com/es-es/wp-content/uploads/sites/16/2019/10/World-Quality-Report-2019-20.pdf

Wang, C., Fayaz, S., Bani, A. Y. A., Ayassrah, A., Mahrous, E., Irshad, M., Ali, Y. A., & Al-razgan, M. (2023). An empirical evaluation of technology acceptance model for Artificial Intelligence in E-commerce. *Heliyon*, *9*. https://doi.org/10.1016/j.heliyon.2023.e18349

Wang, J., Li, Y., & Rao, h. R. (2016). Overconfidence in Phishing Email Detection. *Journal of the Association for Information Systems*, *17*(11), 759–783.

Wang, L., Sun, Y., & Luo, X. R. (2022). Game Affordance, Gamer Orientation, and In-Game Purchases: A Hedonic-Instrumental Framework. *Information Systems Journal*, *32*(6), 1097–1125.

Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1–37. https://doi.org/10.2478/jagi-2019-0002

Wang, X., & Ji, X. (2020). Sample Size Estimation in Clinical Research From Randomized Controlled Trials to Observational Studies. *Chest*, *158*(1), S12–S20.

Wani, T. A., & Ali, S. W. (2015). Innovation Difusion Theory Review & Scope in the Study of Adoption of Smartphones in India. *Journal of General Management Research*, *3*, 101–118.

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2).

Wheatley, K. K., & Wilemon, D. (1999). From Emerging Technology to Competitive Advantage. *PICMET'99: Portland International Conference on Management of Engineering and Technology. Proceedings Vol-1: Book of Summaries (IEEE Cat. No. 99CH36310)*, 35–41.

Wiener, M., Cram, W., & Benlian, A. (2021). Algorithmic Control and Gig Workers: A Legitimacy Perspective of Uber Drivers. *European Journal of Information Systems*, 1–23.

Wilson, V., Srite, M., & Loiacono, E. (2021). The Effects of Item Ordering on Reproducibility in Information Systems Online Survey Research. *Communications of the Association for Information Systems*, *49*, 760–799.

Wulf, J., Mettler, T., & Brenner, W. (2017). Using a Digital Services Capability Model to Assess Readiness for Using a Digital Services Capability Model to Assess Readiness for the Digital Consumer the Digital Consumer. In *MIS Quarterly Executive* (Vol. 16).

Xing, Y., Gong, Y. Z., Wang, Y. W., & Zhang, X. Z. (2014). Path-Wise Test Data Generation Based on Heuristic Look-Ahead Methods. *Mathematical Problems in Engineering*, 1–19.

Xu, N., & Wang, K. (2021). Adopting robot lawyer? The extending artificial intelligence robot lawyer technology acceptance model for legal industry by an exploratory study. *Journal OfManagement & Organization*, *27*, 867–885. https://doi.org/10.1017/jmo.2018.81

Yang, C., Chen, Y., Gong, Q., He, X., Xiao, Y., Huang, Y., & Fu, X. (2018). Understanding the Behavioral Differences Between American and German Users: A Data-Driven Study. *Big Data Mining and Analytics*, *1*(4), 284–296.

Yee, C., Yip, K., & Seng, C. (2023). Computers and Education : Artificial Intelligence Exploring the role of intrinsic motivation in ChatGPT adoption to support active learning : An extension of the technology acceptance model. *Computers and Education: Artificial Intelligence*, *5*. https://doi.org/10.1016/j.caeai.2023.100178

Yin, R. K. (2014). *Case Study Research: Design and Methods*. SAGE.

Zhang, H., Bai, X., & Ma, Z. (2022). Consumer Reactions to AI Design: Exploring Consumer

Willingness to Pay For Ai-Designed Products. *Psychology and Marketing*, *39*(11), 2171–2183.

Zhang, M., Sarker, S., & Sarker, S. (2013). Drivers and Export Performance Impacts of It Capability in 'Born-Global' Firms: A Cross-National Study. *Information Systems Journal*, *23*(5), 419–443.

Zhang, X., Ming, X., Liu, Z., Yin, D., Chen, Z., & Chang, Y. (2019). A Reference Framework and Overall Planning of Industrial Artificial Intelligence (I-AI) For New Application Scenarios. *The International Journal of Advanced Manufacturing Technology*, *101*(9), 2367–2389.

Zhao, Y., Hu, Y., & Gong, J. (2021). Research on International Standardization of Software Quality and Software Testing. *2021 IEEE/ACIS 20th International Fall Conference on Computer and Information Science (ICIS Fall)*, 56–62.

Zhou, J., Fang, Y., & Grover, V. (2022). Managing Collective Enterprise Information Systems Compliance: A Social and Performance Management Context Perspective. *MIS Quarterly*, *46*(1), 71–99.

Zhou, Z., Jin, X. L., Fang, Y., & Vogel, D. (2015). Toward a Theory of Perceived Benefits, Affective Commitment, and Continuance Intention in Social Virtual Worlds: Cultural Values (Indulgence and Individualism) Matter. *European Journal of Information Systems*, *24*, 247–261.

Zöll, A., Eitle, V., & Buxmann, P. (2022). Machine Learning Adoption Based on the TOE Framework: A Quantitative Study. *PACIS 2022 Proceedings*, 1–17.