Human-AI Interaction – Investigating the Impact on Individuals and Organizations

Vom Fachbereich Rechts- und Wirtschaftswissenschaften
der Technischen Universität Darmstadt

genehmigte

Dissertation

von

Felix Peters, M.Sc.
geboren in Bad Soden am Taunus

zur Erlangung des akademischen Grades
Doctor rerum politicarum (Dr. rer. pol.)

Erstgutachter:	Prof. Dr. Peter Buxmann
Zweitgutachter:	Prof. Dr. Dr. Ruth Stock-Homburg
Darmstadt	2022

Peters, Felix: Human-AI Interaction – Investigating the Impact on Individuals and Organizations

Darmstadt, Technische Universität Darmstadt

Dissertation veröffentlicht auf TUprints im Jahr 2023
Tag der mündlichen Prüfung: 01.12.2022

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.

Felix Peters

Darmstadt, 16.09.2022

Abstract

Artificial intelligence (AI) has become increasingly prevalent in consumer and business applications, equally affecting individuals and organizations. The emergence of AI-enabled systems, i.e., systems harnessing AI capabilities that are powered by machine learning (ML), is primarily driven by three technological trends and innovations: increased use of cloud computing allowing large-scale data collection, the development of specialized hardware, and the availability of software tools for developing AI-enabled systems. However, recent research has mainly focused on technological innovations, largely neglecting the interaction between humans and AI-enabled systems. Compared to previous technologies, AI-enabled systems possess some unique characteristics that make the design of human-AI interaction (HAI) particularly challenging. Examples of such challenges include the probabilistic nature of AI-enabled systems due to their dependence on statistical patterns identified in data and their ability to take over predictive tasks previously reserved for humans. Thus, it is widely agreed that existing guidelines for human-computer interaction (HCI) need to be extended to maximize the potential of this groundbreaking technology. This thesis attempts to tackle this research gap by examining both individual-level and organizational-level impacts of increasing HAI.

Regarding the impact of HAI on individuals, two widely discussed issues are how the opacity of complex AI-enabled systems affects the user interaction and how the increasing deployment of AI-enabled systems affects performance on specific tasks. Consequently, papers A and B of this cumulative thesis address these issues.

Paper A addresses the lack of user-centric research in the field of explainable AI (XAI), which is concerned with making AI-enabled systems more transparent for end-users. It is investigated how individuals perceive explainability features of AI-enabled systems, i.e., features which aim to enhance transparency. To answer this research question, an online lab experiment with a subsequent survey is conducted in the context of credit scoring. The contributions of this study are two-fold. First, based on the experiment, it can be observed that individuals positively perceive explainability features and have a significant willingness to pay for them. Second, the theoretical model for explaining the purchase decision shows that increased perceived transparency leads to increased user trust and a more positive evaluation of the AI-enabled system.

Paper B aims to identify task and technology characteristics that determine the fit between an individual's tasks and an AI-enabled system, as this is commonly believed to be the main driver for system utilization and individual performance. Based on a qualitative research approach in the form of expert interviews, AI-specific factors for task and technology characteristics, as well as the task-technology fit, are developed. The resulting theoretical model enables empirical research to investigate the relationship between task-technology fit and individual performance and can also be applied by practitioners to evaluate use cases of AI-enabled system deployment.

While the first part of this thesis discusses individual-level impacts of increasing HAI, the second part is concerned with organizational-level impacts. Papers C and D address how the increasing use of AI-enabled systems within organizations affect organizational justice, i.e., the fairness of decision-making processes, and organizational learning, i.e., the accumulation and dissemination of knowledge.

Paper C addresses the issue of organizational justice, as AI-enabled systems are increasingly supporting decision-making tasks that humans previously conducted on their own. In detail, the study examines the effects of deploying an AI-enabled system in the candidate selection phase of the recruiting process. Through an online lab experiment with recruiters from multinational companies, it is shown that the introduction of so-called CV recommender systems, i.e., systems that identify suitable candidates for a given job, positively influences the procedural justice of the recruiting process. More specifically, the objectivity and consistency of the candidate selection process are strengthened, which constitute two essential components of procedural justice.

Paper D examines how the increasing use of AI-enabled systems influences organizational learning processes. The study derives propositions from conducting a series of agent-based simulations. It is found that AI-enabled systems can take over explorative tasks, which enables organizations to counter the longstanding issue of learning myopia, i.e., the human tendency to favor exploitation over exploration. Moreover, it is shown that the ongoing reconfiguration of deployed AI-enabled systems represents an essential activity for organizations aiming to leverage their full potential. Finally, the results suggest that knowledge created by AI-enabled systems can be particularly beneficial for organizations in turbulent environments.

Zusammenfassung

Künstliche Intelligenz (KI) hat sich in Verbraucher- und Geschäftsanwendungen zunehmend durchgesetzt und beeinflusst Einzelpersonen und Organisationen gleichermaßen. Das Aufkommen von KI-basierten Systemen, d. h. von Systemen, die KI-Fähigkeiten nutzen und auf maschinellem Lernen (ML) basieren, wird in erster Linie durch drei wichtige technologische Trends und Innovationen vorangetrieben: die zunehmende Nutzung von Cloud Computing, die eine groß angelegte Datenerfassung ermöglicht, sowie durch die Entwicklung von spezieller Hardware und Software-Tools für die Entwicklung von KI-basierten Systemen. Die jüngste Forschung hat sich jedoch hauptsächlich auf technologische Innovationen konzentriert und die Interaktion zwischen Menschen und KI-basierten Systemen weitgehend vernachlässigt. Im Vergleich zu früheren Technologien weisen KI-basierte Systeme einige einzigartige Merkmale auf, die die Gestaltung der Mensch-KI-Interaktion (HAI) zu einer besonderen Herausforderung machen. Beispiele für solche Herausforderungen sind die probabilistische Natur KI-basierter Systeme aufgrund ihrer Abhängigkeit von statistischen Mustern, die in Daten identifiziert werden, und ihre Fähigkeit, prädiktive Aufgaben zu übernehmen, die bisher dem Menschen vorbehalten waren. Es besteht daher weitgehend Einigkeit darüber, dass die bestehenden Leitlinien für die Mensch-Computer-Interaktion (HCI) erweitert werden müssen, um das Potenzial dieser bahnbrechenden Technologie voll auszuschöpfen. In dieser Arbeit wird versucht, diese Forschungslücke zu schließen, indem sowohl die Auswirkungen auf individueller als auch auf organisatorischer Ebene untersucht werden, die sich aus der Zunahme von HAI ergeben.

Was die Auswirkungen von KI auf den Einzelnen betrifft, so werden insbesondere zwei Fragen häufig diskutiert: (1) Wie wirkt sich die Undurchsichtigkeit komplexer KI-basierter Systeme auf die Benutzerinteraktion aus und (2) wie wirkt sich der zunehmende Einsatz von KI-basierten Systemen auf die individuelle Leistung bei der Erfüllung bestimmter Aufgaben aus? Daher befassen sich die Studien A und B dieser kumulativen Dissertation mit diesen Fragen.

Studie A befasst sich mit dem Mangel an nutzerzentrierter Forschung auf dem Gebiet der erklärbaren KI, bei der es darum geht, KI-basierte Systeme für Endnutzer transparenter zu machen. Es wird untersucht, wie Individuen Erklärungsfunktionen von KI-basierten Systemen

wahrnehmen, d.h. Funktionen, die darauf abzielen, die Transparenz zu erhöhen. Zur Beantwortung dieser Forschungsfrage wird ein Online-Laborexperiment mit anschließender Befragung im Kontext der Kreditwürdigkeitsprüfung durchgeführt. Diese Studie leistet einen zweifachen Beitrag. Erstens lässt sich anhand des Experiments feststellen, dass Individuen Erklärungsfunktionen positiv wahrnehmen und eine signifikante Zahlungsbereitschaft dafür haben. Zweitens zeigt das theoretische Modell zur Erklärung der Kaufentscheidung, dass eine erhöhte wahrgenommene Transparenz zu einem erhöhten Vertrauen der Nutzer und einer positiveren Bewertung des KI-basierten Systems führt.

Studie B zielt darauf ab, Aufgaben- und Technologiecharakteristika zu identifizieren, die die Passgenauigkeit zwischen den Aufgaben eines Individuums und einem KI-basierten System bestimmen, da dies gemeinhin als Haupttreiber für die Systemnutzung und die individuelle Leistung angesehen wird. Auf der Grundlage eines qualitativen Forschungsansatzes in Form von Experteninterviews werden KI-spezifische Faktoren für Aufgaben- und Technologiemerkmale sowie den Task-Technology Fit entwickelt. Das daraus resultierende theoretische Modell ermöglicht empirische Forschung, die den Zusammenhang zwischen Task-Technology-Fit und individueller Leistung untersucht, und kann auch von Praktikern angewendet werden, um potenzielle Anwendungsfälle des Einsatzes KI-basierter Systeme zu evaluieren.

Während der erste Teil dieser Arbeit die Auswirkungen der zunehmenden KI-Mensch-Interaktion auf individueller Ebene erörtert, befasst sich der zweite Teil mit den Auswirkungen auf organisatorischer Ebene. Die Beiträge C und D befassen sich mit der Frage, wie sich der zunehmende Einsatz KI-basierter Systeme in Organisationen auf die organisatorische Gerechtigkeit, d.h. die Fairness von Entscheidungsprozessen, und das organisatorische Lernen, d.h. die Anhäufung und Verbreitung von Wissen, auswirkt.

Studie C befasst sich mit der Problematik der organisatorischen Gerechtigkeit, die sich aus der Tatsache ergibt, dass KI-basierte Systeme zunehmend Aufgaben übernehmen, die früher von Menschen ausgeführt wurden. Im Einzelnen untersucht die Studie die Auswirkungen des Einsatzes eines KI-basierten Systems in der Phase der Bewerberauswahl im Rekrutierungsprozess. Anhand eines Online-Laborexperiments mit Personalverantwortlichen multinationaler Unternehmen wird gezeigt, dass die Einführung sogenannter CV-Recommender-Systeme, d.h. Systeme, die geeignete Kandidaten für eine bestimmte Stelle identifizieren, die Verfahrensgerechtigkeit des Rekrutierungsprozesses positiv beeinflusst.

Insbesondere werden die Objektivität und die Konsistenz des Bewerberauswahlprozesses gestärkt, welche zwei wesentliche Komponenten der Verfahrensgerechtigkeit darstellen.

Studie D untersucht, wie der zunehmende Einsatz von KI-basierten Systemen organisatorische Lernprozesse beeinflusst. Die Studie leitet Thesen aus der Durchführung einer Reihe von agentenbasierten Simulationen ab. Es wird festgestellt, dass KI-basierte Systeme explorative Aufgaben übernehmen können, was es Organisationen ermöglicht, dem seit langem bestehenden Problem der Lernmyopie entgegenzuwirken, d.h. der menschlichen Tendenz, die Ausschöpfung bestehenden Wissens dem Erkunden neuen Wissens vorzuziehen. Darüber hinaus wird gezeigt, dass die fortlaufende Rekonfiguration von KI-basierten Systemen eine wesentliche Aktivität für Unternehmen darstellt, die deren volles Potenzial ausschöpfen wollen. Schließlich deuten die Ergebnisse darauf hin, dass das von KI-basierten Systemen geschaffene Wissen für Organisationen in turbulenten Umgebungen besonders vorteilhaft sein kann.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
ASA	Attraction-Selection-Attrition
AVE	Average Variances Extracted
B2B	Business-to-Business
B2C	Business-to-Consumer
CDSS	Clinical Decision Support System
CFA	Confirmatory Factor Analysis
CIO	Chief Information Officer
CR	Composite Reliability
CS	Computer Science
CV	Curriculum Vitae
DSS	Decision Support System
EFA	Exploratory Factor Analysis
EU	European Union
GDPR	General Data Protection Regulation
GSS	Group Support System
HAI	Human-AI Interaction
HCI	Human-Computer Interaction
HRM	Human Resource Management
IS	Information Systems
IT	Information Technology
KSA	Knowledge, Skills, Abilities
LDA	Latent Dirichlet Allocation
LSA	Latent Semantic Analysis
ML	Machine Learning
NYOP	Name-Your-Own-Price
OL	Organizational Learning
P-E	Person-Environment

P-G	Person-Group
P-J	Person-Job
P-O	Person-Organization
P-S	Person-Supervisor
P-V	Person-Vocation
R&D	Research and Development
TAM	Technology Acceptance Model
TPB	Theory of Planned Behavior
TTF	Task-Technology Fit
UI	User Interface
UX	User Experience
VIF	Variable Inflaction Factors
WTP	Willingness to Pay
XAI	Explainable Artificial Intelligence

1 Introduction

1.1 Motivation

In recent years, artificial intelligence (AI) has seen increasing use in consumer and organizational contexts. On the consumer side, AI finds its way into digital assistants such as chatbots (e.g., Maedche et al., 2019), improves real-time recommender systems in domains like travel (e.g., Bernardi et al., 2019) or entertainment (e.g., Wei et al., 2017), and powers modern navigation systems such as Google Maps (Derrow-Pinion et al., 2021). Within organizations, AI is set to revolutionize processes such as drug discovery (e.g., Fleming, 2018), manufacturing (e.g., J. Wang et al., 2018), and marketing (e.g., Huang & Rust, 2021). Because the increasing popularity of AI has led to it being used as a buzzword and an umbrella term for many technologies, it is essential to define the research subject for this thesis clearly. First of all, this thesis examines AI-enabled systems, i.e., systems with features harnessing AI capabilities directly exposed to the end-user (Amershi, Weld, et al., 2019). Here, the AI capabilities are realized using machine learning (ML). In its essence, ML, often touted as a "general-purpose technology" (Brynjolfsson & Mitchell, 2017), enables rules to be learned from data in the form of models using learning algorithms (e.g., neural networks, decision trees) instead of encoding them manually, as is traditionally done in software engineering (T. Mitchell, 1997). The resulting models can make predictions on new data samples, such as classifications or forecasts (Brynjolfsson & Mitchell, 2017). ML has become the de-facto standard for realizing AI capabilities, primarily due to three recent trends. First, the increasing use of cloud infrastructure allows large-scale data collection, allowing ML models to be trained on larger and larger datasets, which generally improves their performance (e.g., Brynjolfsson & Mitchell, 2017; Rai et al., 2019). Second, the lowered cost for computing power and development of specialized hardware allows organizations to train AI more efficiently (e.g., Rai et al., 2019; Raisch & Krakowski, 2020). Third, software for developing AI-enabled systems has become more easily accessible for developers, especially in the form of open-source frameworks (e.g., TensorFlow, PyTorch), cloud tools (e.g., Google Cloud AI, AWS SageMaker), and pre-trained ML models (e.g., HuggingFace, PyTorch Image Models) (Rai et al., 2019; Raisch & Krakowski, 2020).

Research on AI-enabled systems has so far mainly been focused on technological innovations. Examples of these kinds of innovations include improved neural network architectures (e.g., Jumper et al., 2021; Vaswani et al., 2017), regularization mechanisms (e.g., Ba et al., 2016; Ioffe & Szegedy, 2015), and data processing algorithms (e.g., Shorten & Khoshgoftaar, 2019; Sohn et al., 2020). A primarily neglected aspect in AI research is the interaction between humans and AI-enabled systems, referred to as human-AI interaction. Due to its unique characteristics, it is widely believed that existing guidelines for human-computer interaction (HCI) (e.g., Höök, 2000; Horvitz, 1999; L. Li & Zhang, 2005; Norman, 1994) might not apply to AI-enabled systems and need to be extended, as postulated in several calls for research in both the IS and HCI research communities (e.g., Maedche et al., 2019; Rai et al., 2019; Schuetz & Venkatesh, 2020; Yang et al., 2020). This thesis aims to tackle this research gap from two perspectives: the impact that increasing human-AI interaction will have on an individual and organizational level.

Regarding the first perspective, i.e., the impact of increasing human-AI interaction on individuals, several unique characteristics of AI-enabled systems require a rethinking of traditional HCI principles. AI-enabled systems are trained by applying statistical learning algorithms (e.g., stochastic gradient descent) to inherently noisy datasets. Thus, the resulting models have a probabilistic nature that might result in unpredictable and inconsistent behavior, especially if the models change over time due to ongoing learning (Ågerfalk, 2020; Amershi, Begel, et al., 2019; Schuetz & Venkatesh, 2020; Yang et al., 2020). For example, AI-enabled systems are prone to making prediction errors, and end-users have to deal with potentially complex outputs, e.g., recommendations in the context of organizational decision-making (Amershi, Weld, et al., 2019; Yang et al., 2020). While a long list of potential issues arise from these unique characteristics, two focus points are discussed in this thesis: transparency and individual performance. A lack of transparency due to the complexity and probabilistic nature of modern ML models is one of the primarily discussed issues in human-AI interaction (e.g., Ågerfalk, 2020; Amershi, Weld, et al., 2019; Maedche et al., 2019). It remains largely unclear how users recognize and perceive the lack of transparency (Maedche et al., 2019; Schuetz & Venkatesh, 2020; Thiebes et al., 2021), particularly in high-stakes contexts such as medicine or finance (Reyes et al., 2020; Rudin, 2019). With regard to individual performance, commonly discussed issues include how support through AI-enabled systems affects human performance on a specific task (Fügener et al., 2021; Grønsund & Aanestad, 2020; Maedche et al., 2019) and how humans and machines should split specific tasks to maximize performance (Fügener et al., 2021; Grønsund & Aanestad, 2020; Schuetz & Venkatesh, 2020).

On an organizational level, the unique characteristics of AI-enabled systems enable large-scale automation and augmentation of processes such as decision-making, especially the speed and scalability to detect valuable patterns in vast amounts of data (Grønsund & Aanestad, 2020; Maedche et al., 2019; Murray et al., 2021; Rai et al., 2019; Raisch & Krakowski, 2020; Schuetz & Venkatesh, 2020). The two focus points of this thesis regarding organizational impacts of increased human-AI interaction are fairness and organizational learning. Fairness is a widely discussed issue when deploying AI-enabled systems. Most research so far has focused on conceptual definitions of fairness in an organizational context (e.g., Mehrabi et al., 2021; Verma & Rubin, 2018). However, it remains unclear whether the use of AI-enabled systems will reinforce or even mitigate existing human biases in decision-making (Maedche et al., 2019; Rai et al., 2019; Raisch & Krakowski, 2020; Thiebes et al., 2021). Regarding organizational learning, researchers agree that knowledge will increasingly be transferred to machines in all possible collaboration modes (e.g., automation, augmentation, assemblage) (e.g., Lyytinen et al., 2021; Raisch & Krakowski, 2020). How this affects organizational learning is an understudied issue, in particular, due to the differences in learning mechanisms and speed between humans and machines (e.g., Lyytinen et al., 2021; Raisch & Krakowski, 2020).

1.2 Structure of the thesis

This thesis contains four papers that examine the impact of increasing human-AI interaction on organizations and individuals. All papers are listed in Table 1. The following paragraphs aim to summarize the content of each paper.

Table 1: List of publications included in this thesis

Paper A	Peters, Felix; Pumplun, Luisa; Buxmann, Peter (2020): **Opening the Black Box: Consumer's Willingness to Pay for Transparency of Intelligent Systems.** In: European Conference on Information Systems (ECIS), A Virtual AIS Conference. VHB-Ranking: B.
Paper B	Sturm, Timo; Peters, Felix (2020): **The Impact of Artificial Intelligence on Individual Performance: Exploring the Fit between Task, Data, and Technology.** In: European Conference on Information Systems (ECIS), A Virtual AIS Conference. VHB-Ranking: B.
Paper C	Eitle, Verena; Peters, Felix; Welsch, Andreas; Buxmann, Peter (2021): **The Impact of CV Recommender Systems on Procedural Justice in Recruiting: An Experiment in Candidate Selection.** In: European Conference on Information Systems (ECIS), A Virtual AIS Conference. VHB-Ranking: B.
Paper D	Sturm, Timo; Gerlach, Jin; Pumplun, Luisa; Mesbah, Neda; Peters, Felix; Tauchert, Christoph; Nan, Ning; Buxmann, Peter (2021): **Coordinating Human and Machine Learning for Effective Organizational Learning.** In: MIS Quarterly, 45 (3), pp. 1581-1602. VHB-Ranking: A+.

Paper A examines the impact of increasing human-AI interaction on individuals, precisely the focus point *transparency*. On a general level, the paper addresses the lack of user-centric research in the field of explainable AI, which aims to develop methods for making AI-enabled systems more transparent. The main research question of paper A is how individuals perceive features that enhance the transparency of AI-enabled systems. Using an online lab experiment and subsequent survey with 223 participants in the context of credit scoring, it is shown that individuals positively perceive such transparency features and even exhibit a significant willingness to pay for them. To explain the mechanisms behind the purchase decision, a theoretical model is developed based on the Theory of Planned Behavior (TPB) (Ajzen, 1991). It can be observed that increased trust due to increased perceived transparency constitutes the most critical factor driving the positive evaluation of transparency features. The study is among the first to empirically examine the effects of deploying transparency features in AI-enabled systems in the social sciences and simultaneously informs practitioners about the relevance of explainable AI.

Paper B also addresses the impacts of increasing human-AI interactions on an individual level but focuses on *performance*. The paper addresses the widely acknowledged research gap of how humans should use AI-enabled systems to maximize individual performance on specific tasks. Following a qualitative research approach, the study aims to establish task and technology characteristics, as well as factors that determine the fit between task and technology with regards to individual performance. Based on twelve expert interviews, an extended version of the task-technology fit (TTF) model (Goodhue & Thompson, 1995) is developed, tailored to the unique characteristics of AI-enabled systems. The model can serve as a basis for future empirical research in this area and enables practitioners to examine the suitability of AI-enabled systems for specific use cases.

While the first two papers cover individual impacts of increasing human-AI interaction, papers C and D are concerned with organizational-level impacts. Paper C addresses the focus point *fairness* and is placed in the context of recruiting. Here, AI-enabled systems are increasingly used to help recruiters cope with vast amounts of available applicant data. The study's central research question is what impact the use of CV recommender systems has on procedural justice in the selection process of candidate recruiting. Among others, procedural justice in recruiting comprises the dimensions of consistency and bias suppression. An online lab experiment was conducted with 74 recruiters from 22 multinational companies, wherein the recruiters were tasked to rank candidates according to their fit for two fictional job postings. The results show that recruiters supported by an AI-enabled CV recommender system created more consistent candidate rankings than unsupported recruiters. In addition, evidence is found that the rankings from supported recruiters are more firmly based on relevant skills and less on personal biases. Thus, the study suggests that AI-enabled systems can positively impact the procedural justice of organizational decision-making.

Paper D extends organizational learning theory in that it softens the assumption that humans are the only learning agents in an organization and thus the sole contributors to organizational knowledge. At its core, the study examines how human and machine learning should be coordinated to maximize organizational knowledge levels. Three main propositions are derived from a series of agent-based simulations. First, the results suggest that well-trained AI-enabled systems can take over explorative tasks, thus allowing humans to learn at their preferred pace. Second, it is found that humans should always be kept in the loop, even for highly automated tasks. This is mainly due to the massive importance of the ongoing reconfiguration of AI-enabled systems. Third, AI-enabled systems can be especially beneficial for organizational learning in turbulent environments. The study contributes to organizational learning theory in that it is the first to account for the unique characteristics of AI-enabled systems.

The papers included in this thesis employ a variety of research methodologies (see Table 2). Papers A, C, and D are quantitative studies; paper B follows a qualitative approach. In detail, papers A and C are based on online lab experiments, paper C on an agent-based simulation, and paper B on expert interviews. Moreover, each paper builds off of different theoretical backgrounds, fitting the respective focus point of the study. The theoretical framework in paper A is based on the theory of planned behavior (Ajzen, 1991), which aims to explain behavioral intention. Paper B extends the task-technology fit theory (Goodhue & Thompson, 1995) to fit the unique characteristics of AI-enabled systems. Paper C uses the attraction-selection-attrition framework (Schneider, 1987) for recruiting in organizations and the organizational justice literature (Greenberg & Colquitt, 2005). Finally, the simulation model in paper D builds on March's model (1991) to examine the effects of increasing human-AI interaction on organizational learning.

Table 2: Outline of research papers

Research paper	Study type	Research methodology	Theoretical background
Paper A: Consumers' Willingness to Pay for Transparency of Intelligent Systems	Quantitative study	Online lab experiment	Theory of Planned Behavior
Paper B: The Impact of Artificial Intelligence on Individual Performance	Qualitative study	Expert interviews	Task-Technology Fit Theory
Paper C: The Impact of CV Recommender Systems on Procedural Justice in Recruiting	Quantitative study	Online lab experiment	Organizational Justice
Paper D: Coordinating Human and Machine Learning for Effective Organizational Learning	Quantitative study	Agent-based simulation	March's Model of Organizational Learning

In addition to the papers included in this thesis (see Table 1), I worked on the following papers, which are either peer-reviewed publications or under review at the time of publishing this thesis:

- Pumplun, Luisa; Fecho, Mariska; Wahl, Nihal; Peters, Felix; Buxmann, Peter (2021): Adoption of Machine Learning Systems for Medical Diagnostics in Clinics: A Qualitative Interview Study. In: Journal of Medical Internet Research, 23(10), e29301, VHB-Ranking: -
- Pumplun, Luisa; Peters, Felix; Gawlitza, Joshua; Buxmann, Peter (2021): Bringing Machine Learning Systems into Medical Practice: A Design Approach to Transparent Machine Learning-Based Clinical Decision Support Systems. Currently under review at Journal of the Association for Information Systems, VHB-Ranking: A

Papers A-D are contained in Chapters 3 to 6 of this thesis.[1] Chapter 2 summarizes the theoretical background regarding human-AI interaction in general and the focus points of this thesis. The theoretical and practical contributions are discussed in Chapter 7, which also concludes this thesis with an outlook on possible future research in this field.

[1] Slight adaptations were conducted to each paper to ensure a consistent paper layout for this thesis. All papers are written in the first-person plural because multiple researchers co-authored them.

2 Theoretical Background

This chapter aims to provide the required theoretical background for the papers included in this thesis. The first section serves as an introduction to AI in general and human-AI interaction in specific. The remaining sections cover the theoretical background for each of this thesis' focus points, namely transparency, individual performance, fairness, and organizational learning.

2.1 Artificial Intelligence

To understand the unique challenges of human-AI interaction, it is necessary to provide some theoretical background regarding modern AI-enabled systems. Therefore, the following two subsections will provide information about modern AI, including its advantages and limitations, and the lifecycle of AI-enabled systems. The remaining subsections will introduce the unique characteristics of AI-enabled systems and state-of-the-art of research in human-AI interaction.

2.1.1 Artificial intelligence and AI-enabled systems

Traditionally, AI research has aimed to create machines that perform cognitive functions usually associated with humans, e.g., "perceiving, reasoning, learning, interacting with the environment, problem-solving, decision-making, and even demonstrating creativity" (Rai et al., 2019, p. iii). Early work on AI was based on rule-based approaches which aimed to mimic human decision-making (Russell & Norvig, 2016). However, this suffered from human's inability to describe their own decision rules, known as Polanyi's paradox (Brynjolfsson & Mitchell, 2017; Fügener et al., 2021). In contrast, modern AI research primarily focuses on statistical machine learning approaches. Here, the machine learns models from data by detecting patterns. The resulting model can then make predictions on new data (Brynjolfsson & Mitchell, 2017). Especially the advent of deep learning, i.e., the training of deep neural networks, has enabled machines to perform tasks at or even above human-level performance. Examples of these tasks include image recognition (e.g., He et al., 2016; Krizhevsky et al., 2012), language modeling (e.g., Howard & Ruder, 2018; Vaswani et al., 2017), playing board and online games (e.g., Berner et al., 2019; Silver et al., 2018), and most recently predicting the three-dimensional structure of proteins (Jumper et al., 2021).

Based on the recent progress, AI-enabled systems have gained more and more traction in organizations of all industries. Due to the high cost of training machine learning models, researchers and practitioners have aimed to develop guidelines for when to rely on AI-enabled systems. In its essence, machine learning enables machines to make predictions based on previously detected patterns in data (Agrawal et al., 2019). Thus, organizations can leverage the speed and scalability of computers for performing tasks that rely on these predictions (Rai et al., 2019). However, tasks need to be clearly defined, meaning that clear metrics exist for providing feedback on task performance and that the input data contains relevant information for performing the task (Brynjolfsson & Mitchell, 2017). In addition, tasks performed by an AI-enabled system should rely on statistical patterns, not on in-depth reasoning (Brynjolfsson & Mitchell, 2017). Modern AI-enabled system's limitations include the inability to perform complex reasoning and logic, the lack of explainability of model predictions, and learning from small datasets (Brynjolfsson & Mitchell, 2017; Rudin, 2019).

2.1.2 Lifecycle of AI-enabled systems

Like most software systems, AI-enabled systems follow an iterative development process. Various process models exist, which commonly include the following phases (e.g., Amershi, Begel, et al., 2019; Sturm, Gerlach, et al., 2021). Initially, the system requirements need to be defined, especially the tasks performed by AI. Data for training the required ML models need to be collected, preprocessed, and labeled (in the case of supervised learning). Based on the created datasets, models can be trained and evaluated before the AI-enabled system can be deployed into a production setting. The deployed models need to be monitored regarding performance (e.g., accuracy) and operational metrics (e.g., latency, throughput, memory usage). As mentioned above, these phases are not followed in a sequential but iterative manner. For example, model evaluation or monitoring might uncover new data requirements (Amershi, Begel, et al., 2019). Apart from being the potential users of AI-enabled systems, humans are involved in all stages of the development process. While domain experts are required for defining the prediction tasks, developers (e.g., data scientists, ML engineers) prepare the data for model training, select and tune the learning algorithms, and evaluate model performance (Amershi, Begel, et al., 2019; Grønsund & Aanestad, 2020; Raisch & Krakowski, 2020). It is also on humans to continually monitor the system and decide on reconfiguration efforts such as retraining models based on new data or improving the data quality (Baylor et al., 2017; Sculley et al., 2015; Sturm, Gerlach, et al., 2021).

2.1.3 Unique characteristics of AI-enabled systems

Given this understanding of AI-enabled systems, two unique characteristics of these systems can be identified, distinguishing them from traditional IT systems.

First, due to their dependence on data and statistical machine learning algorithms, AI-enabled systems are inherently probabilistic (Ågerfalk, 2020; Schuetz & Venkatesh, 2020; Yang et al., 2020). Consequently, these systems are prone to making prediction errors, especially in cases not sufficiently represented in the training data (Amershi, Weld, et al., 2019; Yang et al., 2020). For example, an AI-enabled system tasked with detecting fraud in credit card transactions (i.e., a binary classification problem) can make two types of errors: (1) predicting a fraudulent transaction as valid (false negative), and (2) predicting a valid transaction as fraudulent (false positive). Another consequence of the probabilistic nature of AI-enabled systems is potential inconsistency in system behavior (Amershi, Weld, et al., 2019; Schuetz & Venkatesh, 2020). As mentioned in the previous subsection, AI-enabled systems need to be continually reconfigured, for example, in the form of retraining on new data. However, this can lead to changes in predictions over time. For the example presented above, reconfiguration could include adding new features to the dataset used to train the fraud detection system. If this feature is highly predictive of fraudulent transactions, several transactions previously classified as valid might now be classified as fraudulent.

Second, the prediction capabilities of AI-enabled systems allow organizations to automate and augment tasks previously reserved for humans (Maedche et al., 2019; Rai et al., 2019; Raisch & Krakowski, 2020). As a result, it is widely agreed that new collaboration modes have to be identified to maximize the potential of AI-enabled systems by combining the strengths of humans and machines (Lyytinen et al., 2021; Raisch & Krakowski, 2020). Moreover, arising issues from the increasing automation of decision-making revolve around the governance of these systems (Rai et al., 2019). For example, it remains unclear whether AI-enabled systems will reinforce or mitigate cognitive biases in human decision-making (Maedche et al., 2019; Rai et al., 2019; Raisch & Krakowski, 2020; Thiebes et al., 2021).

2.1.4 State-of-the-art in human-AI interaction research

The unique characteristics of AI-enabled systems challenge existing guidelines in the field of human-computer interaction (HCI) (e.g., Höök, 2000; Horvitz, 1999; L. Li & Zhang, 2005; Norman, 1994), which have led to calls for more targeted research on human-AI interaction (e.g., Maedche et al., 2019; Rai et al., 2019; Schuetz & Venkatesh, 2020; Yang et al., 2020). Recent research on human-AI interaction is presented in this subsection, divided into conceptual and empirical work, and listed in chronological order.

Regarding conceptual research, the following studies are relevant to the topics discussed in this thesis. Rzepka and Berger (2018) conduct a systematic review of IS literature on human-AI interaction and create a research framework, adopting the framework of Li and Zhang (2005). The main limitation of this study is the sole focus on the IS discipline, neglecting the interdisciplinary nature of this research area. Amershi et al. (2019) propose high-level design guidelines for human-AI interaction based on a literature review, user studies, and expert interviews. However, the guidelines are not empirically validated, e.g., by measuring desired outputs such as trust. Thus, the relevance of single guidelines and potential trade-offs between them remain unclear. Yang et al. (2020) draw on the experience of designers and UX experts to identify the two key challenges of designing UI for AI-enabled systems: capability uncertainty (i.e., functionality, performance, kinds of produced errors) and output complexity (from simple binary classification to complex recommendations). Furthermore, they derive a framework for addressing these issues in design processes. Lyytinen et al. (2021) assemble a research agenda for meta-human systems, i.e., systems in which both humans and machines learn (including, but not limited to, AI-enabled systems, as discussed in this thesis). They propose research questions related to the organizational functions of delegating, monitoring, cultivating, and reflecting.

Additionally, empirical studies have studied contexts in which humans and AI-enabled systems interact in various ways. For example, Yang et al. (2019) develop a clinical decision support system (CDSS) for artificial heart implementation, including prognostic prediction from an AI-enabled system. In experimental studies with physicians, they investigate how the predictions should be presented to achieve user acceptance and trust. They find that the correct level of remarkableness needs to be identified to not distract physicians from the task at hand by solely focusing on the characteristics of the AI-enabled system. Grønsund & Aanestad (2020) examine how human-in-the-loop configurations can be designed when introducing AI-enabled systems. Based on a qualitative case study in a financial services and brokering firm involved in global maritime trade, they describe how workflows in this specific organization were reconfigured to maximize performance. They propose an emergent human-in-the-loop configuration for algorithmic decision-making involving multiple feedback cycles between humans and the AI-enabled system. Also, in the context of financial decision-making, Sturm et al. (2021) study how humans and AI-enabled systems should interact when trading securities. They find that the best trading performance is achieved when humans and machines learn from each other. Thus, they stress the importance of including feedback cycles to combine the strengths of humans and AI-enabled systems. Fügener et al. (2021) examine whether delegation between humans and

AI outperforms humans or AI working alone and which factors limit human delegation. Using experimental studies within the context of image recognition, they find that the best performance is achieved when humans and AI collaborate, primarily when the AI handles most of the decisions and delegates cases to the human, where the AI is uncertain.

2.2 Transparency and Explainable AI

Transparency research has been a theme in IS research for a while. Early work on this topic focused on Knowledge-Based Systems, which were usually not based on statistical but logical ML approaches (e.g., Dhaliwal & Benbasat, 1996; Gregor & Benbasat, 1999; Ye & Johnson, 1995). For the modern AI-enabled systems discussed in this thesis, transparency research primarily takes place under the umbrella term explainable AI (XAI). XAI is a research field at the intersection of computer science (CS), HCI, and social sciences. It is concerned with the development of AI-enabled systems which are understandable by humans (Barredo Arrieta et al., 2020). The primary interface between AI-enabled systems and humans is explanations, which are incorporated with the primary goal of increasing system transparency (T. Miller, 2019). In the following, goals and influencing factors for designing explanations are presented, along with different explanation types.

Three main goals are commonly pursued by incorporating explanations into AI-enabled systems. First, developers such as data scientists or ML engineers can better understand ML model behavior, offering them the opportunity of detecting errors and improving performance (U. Bhatt et al., 2020; Doshi-Velez & Kim, 2017; Lipton, 2016). Second, explanations can be deployed for compliance and regulatory reasons (U. Bhatt et al., 2020; Cheng et al., 2019; Reyes et al., 2020). For instance, the EU's General Data Protection Regulation (GDPR) includes a "right to explanation" for outputs of AI-enabled systems which concern the personal attributes of consumers (Goodman & Flaxman, 2017). In organizational contexts, explanations might be used to pass internal or external audits of AI-enabled systems (U. Bhatt et al., 2020; Reyes et al., 2020). Third, explanations can be used to enhance the interaction between an AI-enabled system and the end-user (Hohman et al., 2019; Liao et al., 2020; D. Wang et al., 2019). However, there is not enough research yet that empirically supports this relationship.

Prior XAI research has also identified influencing factors for designing explanations in specific application scenarios. These include the ML model and data type used to train the AI-enabled system, user characteristics, and the decision context (Liao et al., 2020; Reyes et al., 2020). Regarding the ML model type, the main distinction is between black-box and white-box models. White-box models have built-in explanations (e.g., rules extracted from a decision tree), whereas black-box models (e.g., deep neural networks) require additional techniques to explain model behavior. The type of training data also influences explanation design. For example, simple if-then rules can be used as explanations for AI-enabled systems dealing with structured (i.e., tabular) data. Presentation formats need to vary for AI-enabled systems that operate on unstructured data such as images or texts. User characteristics that influence the explanation design include AI and domain knowledge and the user's general attitude towards AI-enabled systems. Regarding the decision context, outcome criticality (i.e., high-stakes vs. low-stakes decisions), time sensitivity, or decision complexity are commonly named essential factors (Liao et al., 2020; Reyes et al., 2020).

The possible content of explanations can broadly be grouped into model explanations, global explanations, and local explanations. Model explanations comprise general information about the AI-enabled system (e.g., version, intended use, limitations, included ML models), conducted validation procedures (e.g., regulatory approvals, certificates, scientific publications), metadata about the utilized datasets for training the ML models (e.g., data scheme, collection, statistics, labeling procedure, preprocessing), and performance evaluation (e.g., metrics on unseen test data, performance trade-offs) (e.g., Diakopoulos, 2016; Gebru et al., 2021; M. Mitchell et al., 2019). Global explanations describe how the AI-enabled system works on a general level. Possible explanations include average feature importance scores, the importance of other human-defined criteria, or visualizations of internal parameter spaces (e.g., Ghorbani et al., 2019; Goldstein et al., 2015; Kim et al., 2018; Olah et al., 2018). In contrast to that, local explanations provide the rationale for single predictions. Concrete forms of explanation type include feature attributions (e.g., Lundberg & Lee, 2017; Ribeiro et al., 2016; Sundararajan et al., 2017), counterfactual explanations (e.g., Fernandez & Provost, 2019; Mothilal et al., 2020; Wachter et al., 2017), uncertainty estimates (e.g., Gal & Ghahramani, 2016; Guo et al., 2017; Maddox et al., 2019), sensitivity analysis (e.g., Koh & Liang, 2017; Kwon et al., 2019; Ribeiro et al., 2018), and example-based explanations (e.g., Biggio & Roli, 2018; Goodfellow et al., 2014; Kim et al., 2016).

Despite the growing literature on XAI, technical limitations still exist explaining the behavior of complex AI-enabled systems (U. Bhatt et al., 2020; Liao et al., 2020). In particular, explanations for AI-enabled systems are typically based on statistical algorithms, meaning they can become very complex or unintuitive to humans and even contain errors (e.g., T. Miller, 2019; Rudin, 2019). While researchers and practitioners agree on the overarching goals of deploying XAI, it remains largely unclear how to design explanations in practice and how end-users will perceive explanations. Technical XAI studies primarily focus on developing particular explanations, neglecting the interaction between end-users and ML systems.

2.3 Individual Performance and Task-Technology Fit

The task-technology fit (TTF) theory is the most widely used theoretical model for studying performance impacts on an individual level due to technology use. The TTF theory was first proposed by Goodhue and Thompson (1995) and is based on the premise that technology must match the task it is supposed to support to be utilized. The authors further claim that utilization is the primary driver of performance impacts. In summary, the TTF theory comprises five primary constructs (see Figure 1): task characteristics, technology characteristics, task-technology fit, utilization, and performance impacts. In the seminal paper, the task-technology fit construct is operationalized with eight factors (quality, locatability, authorization, compatibility, ease of use, production timeliness, systems reliability, relationship with users). Utilization can be measured by examining usage frequency, and performance impacts can be observed as (perceived) increases in effectiveness, efficiency, or quality when performing a specific task. The original theory has since been extended and applied in various contexts. The following paragraph summarizes IS research related to the TTF theory.

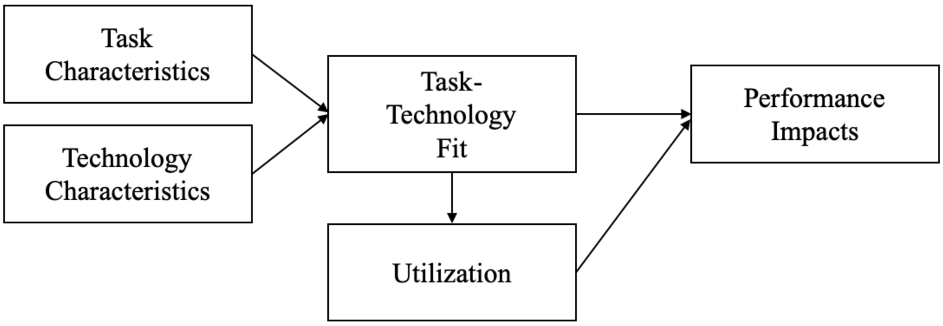

Figure 1: Task-technology fit (TTF) theory

A large body of research transferred the TTF theory to the team level, examining the effects of deploying group support systems (GSS). The first such study was conducted by Zigurs and Buckland (1998). They developed TTF-based models to explain group performance for five different task types (simple, problem, decision, judgment, and fuzzy tasks). In further research, TTF was found to positively impact group-level performance metrics such as decision quality, satisfaction, team commitment, and group cohesion (Dennis et al., 2001; Maruping & Agarwal, 2004). Another stream of research examines how TTF relates to appropriation effects, i.e., how teams adapt technology to their needs over time. Fuller and Dennis (2009) showed that TTF is the main predictor for team performance in the short term. However, appropriation effects are a better predictor in the long term as they reflect a team's ability to adapt a given technology for the task at hand.

On an individual level, the TTF theory was applied in various contexts. When examining performance impacts of mobile IS (e.g., digital assistants, mobile e-commerce), the theory was extended with a context construct to account for the unique characteristics of mobile IS, operationalized by factors such as degree of distraction, connection quality, and user mobility (Gebauer et al., 2010; Gebauer & Ginsburg, 2009). Additionally, TTF was used to study outcomes of deploying systems with similar capabilities compared to AI-enabled systems, e.g., data analytics systems and decision support systems (DSS). For example, Wongpuninwatana et al. (2000) built a theoretical model based on the TTF theory to study individual performance for an auditing task supported by an expert system, i.e., a form of DSS. The TTF theory was also combined with the Technology Acceptance Model (TAM) to examine the intention to use intelligent agents in tasks related to web-based auction processes, e.g., price negotiation and item acquisition (Chang, 2008). Here, the TTF construct was shown to be a predictor for some of the primary TAM constructs (e.g., perceived usefulness, perceived ease of use). In the field of data analytics, the TTF theory was used to study user satisfaction with data in general (Karimi et al., 2004), the relationship between the use of data analytics systems and organizational agility (Ghasemaghaei et al., 2017), and DSS support for specific tasks such as insolvency legislation (Parkes, 2013).

Although a large body of research exists examining the influence of technology on individual performance, the findings cannot simply be transferred to modern AI-enabled systems due to their unique characteristics. In contrast to deterministic traditional IT systems, AI-enabled systems are based on statistical patterns found in data, potentially leading to error-prone or inconsistent behavior. The reliance on data is not accounted for in the classical TTF theory constructs. Thus, the theory needs to be extended to explain the performance impacts of deploying AI-enabled systems.

2.4 Fairness and Organizational Justice

As AI-enabled systems are increasingly deployed to support organizational tasks, a growing body of literature has evolved around the notion of algorithmic fairness. Here, fairness is defined as the "absence of any prejudice or favoritism toward an individual or group based on their inherent or acquired characteristics" (Mehrabi et al., 2021, p. 1). In the following paragraphs, the notions of fairness and justice are introduced first from a technical and then from an organizational perspective.

The technical literature on fairness is primarily concerned with formal definitions for measuring fairness, sources of bias in AI-enabled systems, and mitigation strategies to overcome these biases. Definitions of fairness can be grouped into three different categories (Mehrabi et al., 2021; Verma & Rubin, 2018). First, definitions focusing on individual fairness ensure that AI-enabled systems drive similar predictions to similar individuals. Second, group fairness definitions aim to treat different population groups equally, for example, measured by performance metrics such as accuracy. Third, definitions related to subgroup fairness combine the previous two approaches, aiming to receive similar performance levels and predictions for a large variety of subpopulations. In practice, it is typically not possible to satisfy all three criteria at once (Kleinberg et al., 2017). Biases can occur in all major components of the lifecycle of AI-enabled systems, especially the training data, learning algorithm, and user interactions. However, the most common source of biases in the training data (Mehrabi et al., 2021; Olteanu et al., 2019; Suresh & Guttag, 2019). While a complete list of potential data biases is beyond the scope of this thesis, examples include measurement biases (e.g., choosing inaccurate proxy variables as labels) or sampling biases (e.g., underrepresentation of specific subpopulations). Aiming to overcome these biases and create fair AI-enabled systems, mitigation strategies can be classified into pre-, in-, and post-processing techniques (d'Alessandro et al., 2017; Mehrabi et al., 2021). Pre-processing techniques try to remove the biases from the training data before applying the learning algorithm to it. In-processing techniques typically target the learning algorithm, for example, by modifying the objective function. Finally, post-processing techniques alter the predictions from an AI-enabled system to satisfy selected fairness metrics.

From an organizational perspective, examining fairness in decision-making processes requires understanding the literature on organizational justice. This body of research is primarily concerned with employees' reactions to unfairness and inequity in an organizational context, commonly distinguishing between distributive and procedural justice (Colquitt, 2001; Greenberg & Colquitt, 2005; Leventhal, 1980). Here, distributive justice refers to how employees perceive the distribution of outcomes (e.g., salaries, rewards) as fair. The notion of distributive fairness thus comprises concepts such as equity theory (Adams, 1965; Cohen, 1987) and equality (Deutsch, 1975). While distributive justice is more concerned with the distribution of outcomes, procedural justice focuses on the perceived fairness of the decision-making processes which lead to these outcomes (Greenberg & Colquitt, 2005). Leventhal (1980) described six factors that influence procedural justice: consistency (i.e., applying a consistent process over time and persons), bias suppression (i.e., neutral and unbiased decision-makers), information accuracy (i.e., no inaccurate information is relied on), correctability (i.e., bad outcomes can be appealed), representativeness (i.e., all affected subgroups are included in the process), and ethicality (i.e., applying general standards of ethics and morality).

All in all, the role of AI-enabled systems regarding fairness in decision-making processes remains largely unclear. On the one hand, AI-enabled systems could help to mitigate social biases, primarily when humans and AI systems work together (Rai et al., 2019; Raisch & Krakowski, 2020); on the other hand, organizations need to ensure that existing stereotypes are not reinforced by including biases in training data and algorithms (Amershi, Weld, et al., 2019; Maedche et al., 2019; Rai et al., 2019; Raisch & Krakowski, 2020). Recent literature in the context of recruiting illustrates this duality. While qualitative studies found that both applicants and recruiters expect the deployment of AI-enabled systems to have a positive impact on procedural justice in the recruiting process (Ochmann & Laumer, 2019; Thielsch et al., 2012), Amazon had to pull an AI-enabled system for prioritizing applications because the system systematically preferred male candidates due to an overrepresentation of these in the training data (Logg, 2019). In summary, more research on the impact of increasing human-AI interaction on fairness in organizational decision-making processes is needed, as postulated in several calls for research (Maedche et al., 2019; Rai et al., 2019; Raisch & Krakowski, 2020).

2.5 Organizational Learning

As one of the critical predictors for firm performance, organizational learning research has a long history (Hatch & Dyer, 2004; Levinthal & March, 1993; March, 1991). Organizational learning can be defined as drawing lessons from historical events and incorporating them into organizational routines which aim to guide behavior by organizational members (Levitt & March, 1988). As a result, knowledge is stored in some form of organizational memory, such as rules, procedures, or technology (e.g., Feldman & Pentland, 2003; Levitt & March, 1988). From a macro perspective, organizational learning needs to be managed by coordinating the activities and interactions between organizational members to achieve optimal learning effectiveness (e.g., Dodgson et al., 2013; Fang et al., 2010; March, 1991; K. D. Miller et al., 2006).

Because the knowledge in an organization results from aggregating the beliefs of individual members of the organization, belief diversity is a crucial predictor for organizational learning effectiveness (March, 1991). How to effectively manage belief diversity has been a theme of organizational learning research, as it is by no means a trivial problem (Lavie et al., 2010; March, 1991). For example, organizations that disseminate apparently correct beliefs quickly reduce belief diversity. This might lead to undesired consequences such as adapting inferior routines (Denrell & March, 2001; Levinthal & March, 1993; March, 1991). Therefore, organizations need to maintain a healthy level of belief diversity to experiment with different ideas before converging to the optimal set of beliefs. Consequently, managing the learning speed of organizational members is one of the primary mechanisms for coordinating organizational learning. Here, organizations need to manage the so-called exploration-exploitation trade-off (e.g., Lavie et al., 2010; March, 1991). While exploration mechanisms are approaches for maintaining the belief diversity among organizational members by experimenting with new ideas and uncertain outcomes, exploitation mechanisms converge beliefs to a de-facto status quo reaping certain, immediate benefits. Managing this trade-off resembles the key activity of coordinating organizational learning. An overemphasis on exploitation to gain short-term benefits can lead to a stagnant organization unable to explore new and promising directions. In contrast, leaning too strongly towards exploration might threaten firm survival in competitive environments, as there is no shared understanding of the organization's key competencies (e.g., Lavie et al., 2010; March, 1991; Raisch et al., 2009).

Previous organizational learning research has examined the effects of IT systems on organizational learning. However, two issues arise when examining how increased human-AI interaction might impact organizational learning effectiveness. First, there is no consensus on whether IT actually facilitates or hinders the creation and dissemination of knowledge in organizations (e.g., Alavi & Leidner, 2001; Robey et al., 2000; Schultze & Leidner, 2002). On the one hand, it might be argued that IT simplifies communication between organizational members, leading to simplified dissemination of knowledge. On the other hand, IT aiming at automating processes might erase knowledge from the organizational memory because it is no longer needed (e.g., Iyengar et al., 2015; Robey et al., 2000; Schultze & Leidner, 2002). Second, most studies assume that IT systems' sole purpose is to support humans (e.g., Iyengar et al., 2015; Kane & Alavi, 2007; Robey et al., 2000). This stands in contrast to the assumption that AI-enabled systems might possess agency, learning and contributing knowledge through autonomous learning (e.g., Ransbotham et al., 2020; Seidel et al., 2019). Existing studies that acknowledge AI-enabled systems' self-learning capabilities are on a conceptual level or limited to specific use cases within organizations (e.g., G. D. Bhatt & Zaveri, 2002; Elofson & Konsynski, 1993; Zhu et al., 1997). Thus, how the increasing interaction between humans and AI-enabled systems impacts organizational learning processes remains largely an open question.

3 Paper A: Consumer's Willingness to Pay for Transparency of Intelligent Systems

Title

Opening the Black Box: Consumer's Willingness to Pay for Transparency of Intelligent Systems

Authors

- Felix Peters, Technical University of Darmstadt, Germany
- Luisa Pumplun, Technical University of Darmstadt, Germany
- Peter Buxmann, Technical University of Darmstadt, Germany

Publication Outlet

Proceedings of the Twenty-Eighth European Conference on Information Systems (ECIS 2020), A Virtual AIS Conference

Abstract

Artificial intelligence (AI) is becoming increasingly popular and intelligent systems are deployed for various use cases. However, as these systems typically rely on complex machine learning methods, they effectively exemplify black boxes. Thus, consumers are usually not informed about inner workings of these systems, e.g., data sources or feature importance. Public and private institutions have already called for fairness and transparency standards regarding intelligent systems. Although researchers develop mechanisms to ensure transparency of intelligent systems, it remains an open question how consumers perceive such transparency features. Consequently, our study examines to what extent consumers are willing to pay for these features, and what the underlying mechanisms of the purchase decision are. To answer these questions, we conduct an experiment and a subsequent survey in the context of credit scoring. We show that consumers exhibit significant willingness to pay for transparency. Furthermore, we observe that increased trust in the intelligent system caused by enhanced perceived transparency is the main driver for positive evaluation of transparency features. Our findings inform practitioners about the relevance of "fair AI" and manifest the importance of transparency research regarding intelligent systems in social sciences.

Keywords

Intelligent Systems, Machine Learning, Transparency, Willingness to Pay

3.1 Introduction

The field of artificial intelligence (AI) has seen a lot of technological breakthroughs in recent years: AI was able to defeat their human counterparts in strategic board (Silver et al., 2017) and online games (Vinyals et al., 2019) and has surpassed human-level performance on tasks like image recognition (He et al., 2016). Moreover, AI is increasingly being used for making predictions in high-stakes situations, e.g., medical diagnosis (Kourou et al., 2015) or credit scoring (Burrell, 2015; Zhou, 2017). Modern AI mainly uses machine learning (ML) to enable information systems to do tasks that we used to think were reserved for humans (Andrews et al., 2018; Brynjolfsson & Mitchell, 2017; Elliot & Andrews, 2017; Rai et al., 2019). These intelligent systems can be distinguished from conventional systems since they are able to learn based on incoming data, adapt their behavior over time without having to be explicitly reprogrammed, derive their results statistically, and thus will make mistakes (Andrews et al., 2018; Brynjolfsson & Mitchell, 2017). High complexity of datasets and ML algorithms leads to what is commonly referred to as black box behavior: the lack of transparency in decision-making processes of intelligent systems. In this context, decisions which are inaccurate or not aligned with ethical standards due to biased incoming data could stay undetected (Kruse et al., 2019; Rudin, 2019). Therefore, demands to open the black box of intelligent systems applied to high-stakes decisions (e.g., credit scoring) are increasingly claimed in research and practice (Burrell, 2015; Chen et al., 2018; Google, 2019; Rudin, 2019; Shook et al., 2018; The Economist, 2018). These requests coincide with the declared aim of the United Nations to support fairness, accountability and transparency of intelligent systems (ITU, 2018).

The transparency issue regarding intelligent systems can be exemplified in the credit scoring context. So far, credit applicants could be sure that decisions about their creditworthiness were based on the evaluation of their credit history and the credit documents submitted. Furthermore, simple linear modeling approaches comprising a restricted number of well-known variables have been employed by financial institutions to guarantee the comprehensibility of the credit scoring process. This way, financial institutions are able to provide specific reasons to the credit applicant why one's credit was granted or denied (Martens et al., 2007). However, an increasing number of financial institutions additionally start to rely on tech companies such as Kreditech, Branch, or FICO to determine the creditworthiness of their applicants (Burrell, 2015). These companies automatically collect applicant's personal data from online sources (e.g., corporate websites) and use complex non-linear models based on ML (F.-C. Li et al., 2009; Zhou, 2017). While these companies help consumers with limited financial history or past financial difficulties (e.g., in developing countries) getting access to credits (e.g., FICO, 2019b), decisions about creditworthiness become less transparent than they were before. While the applicant has previously been granted the opportunity of clear reasoning, intelligent systems are limited in their comprehensibility and can hardly be validated to see if they discriminate applicants, e.g., due to gender, religion, or national origin (Shook et al., 2018).

To counteract this problem, a vast amount of research has been conducted over the past few years concerning the technical realization of transparency features for ML algorithms, e.g., neural networks (e.g., Diakopoulos, 2016; Doshi-Velez & Kim, 2017; Kim et al., 2018). In this context, more and more providers of intelligent systems for credit scoring are also striving to make their products more transparent (FICO, 2019a). In contrast, information systems (IS) research has only recently begun to analyze the added value derived by transparency features and its effect on user’s behavior in order to contribute to a responsible design of intelligent systems. In this context, Rzepka and Berger (2018) claim that a transparent system appearance with appropriate explanations shows positive impact on user’s adoption decision. However, it is not yet known how consumers from the general public, who will increasingly depend on decisions from intelligent systems, assess enhanced transparency of intelligent systems since IS research on this topic is still in its infancy. In order to contribute to a more holistic picture of consumer’s assessment of transparency features for intelligent systems, we examine not only their attitude towards these features but also go a step further and investigate their willingness to pay (WTP). This approach aims to demonstrate the interest of an emancipated consumer in more transparency while at the same time creating an incentive for companies to invest resources in improving the explainability of intelligent systems and offer them to customers. Thus, we pose the following research questions:

(RQ1) To what extent are consumers willing to pay for transparency features in the context of intelligent systems?

(RQ2) What are the underlying mechanisms of the purchase decision for transparency features of intelligent systems?

Aiming for an understanding of the value consumers assign to enhanced transparency, this study is the first to explore WTP for transparency features of intelligent systems. Therefore, we run an experiment in the credit scoring context, analyzing whether and why consumers would be willing to pay for features that explain rationale about how their creditworthiness was predicted by an intelligent system. Furthermore, the underlying effects of purchasing behavior are examined by conducting a subsequent survey with the 195 experiment participants. The credit scoring context is chosen because it has been used to stress the importance of transparency of intelligent systems repeatedly in prior research (Burrell, 2015; Kruse et al., 2019; Shook et al., 2018). We show that users exhibit a significant WTP for transparency features and that an increase in trust towards the intelligent credit scoring system is the main reason why consumers positively evaluate such features. As a result, we find a profound interest of consumers for fair and transparent AI, which is not only reflected in their attitude towards transparency features but is actually characterized by a significant WTP. In this regard, the positive evaluation of transparency features enables practitioners to assess their economical relevance for intelligent systems.

Thus, the remainder of this paper is structured as follows: In the beginning, an overview of related work concerning intelligent systems and possible transparency features is provided in order to mark out the research area. Here, we consider prior work from both information systems and computer science. Subsequently, an initial research model is derived and the quantitative research approach, including the experiment as well as the survey design, is presented. After introducing our sample, the results of the study are analyzed. In our discussion, we shed further light onto our contributions, limitations and opportunities for future research. Finally, we conclude our paper by summarizing the investigated problem, our research approach and findings.

3.2 Related Work

3.2.1 Intelligent Systems

AI as a research area covers aspects of mathematics, economics, computer engineering, cybernetics and linguistics, and is concerned with the development and understanding of intelligent systems (Russell & Norvig, 2016). Intelligent systems such as speech-based assistants, search engines or credit scoring systems are already used by private persons in everyday life (Burrell, 2015; von Krogh, 2018). Not only consumers but also organizations are increasingly interested in intelligent systems, for example to recommend financial products, process transactions or schedule complex logistics (Bamberger, 2018). Intelligent systems are generally classified into weak or strong AI, depending on the scope of their task. While the research area of strong AI aims to create a general human-like intelligence, weak AI refers to the solving of a specific problem (Kurzweil, 2005). So far, all intelligent systems can be assigned to weak AI. Here, ML is increasingly used as the underlying technology, frequently combined with more classical approaches such as searching or planning algorithms (Russell & Norvig, 2016; Silver et al., 2017). Intelligent systems based on ML are software programs that increase their performance with respect to a particular task by gaining more experience (T. Mitchell, 1997). In this context, experience corresponds to the data (e.g., images, numbers, texts) that is used for training (Crowston & Bolici, 2019; von Krogh, 2018). Intelligent systems thus differ from conventional systems as they do not rely on prespecified instructions of developers, but develop internal representations based on patterns observed in the training data that are used for making predictions (Burrell, 2015). As a result, intelligent systems have the ability to learn from user's behavior, deduce autonomously and react to their environment; characteristics that are usually reserved to humans (Rai et al., 2019). Interacting with intelligent systems can therefore lead users to perceive humaneness and be threatened by the system (Rzepka & Berger, 2018). Furthermore, by learning progressively to provide users with predictions, intelligent systems can exhibit black box behavior, especially when relying on complex statistical ML methods (Adadi & Berrada, 2018). Therefore, the attitude of users towards intelligent systems is significantly determined by the transparency of their underlying process for making predictions. In the following, we provide an overview of both computer science and information systems perspectives on transparency.

3.2.2 Computer Science Perspective on Transparency

In computer science (CS) literature, transparency research can be grouped under the larger research stream of intelligent systems alignment. Research in this area aims to ensure alignment between intelligent systems and human interests, mainly by developing mechanisms to facilitate algorithmic accountability (Chakraborti et al., 2019; Diakopoulos, 2016). In general, alignment approaches can be categorized into cooperative and adversarial approaches. Cooperative approaches intend to facilitate alignment through understanding system behavior, in contrast to adversarial approaches that are more focused on avoiding misalignment, e.g., by introducing safety measures (Chakraborti et al., 2019). Transparency mechanisms are commonly categorized as a cooperative approach to create intelligent system alignment. Moreover, it is widely assumed that demand for transparency arises from a mismatch between the formal objectives of intelligent systems and the real-world costs occurring in a deployment setting (Doshi-Velez & Kim, 2017; Lipton, 2016). Studies in CS literature often presume that transparency serves as a proxy for adoption drivers such as perceived trust, fairness and user satisfaction (Diakopoulos, 2016; Doshi-Velez & Kim, 2017). However, these relationships are typically not empirically confirmed. Other theoretical work in technical literature also covers limitations of the transparency ideal in intelligent systems. According to these studies, technical limitations especially occur in systems whose inherent complexity is naturally hard to grasp for humans (e.g., deep learning systems). Moreover, system predictions are commonly state- and thus time-dependent, requiring transparency features to store previous versions of the intelligent system in order to explain predictions at an earlier point in time (Ananny & Crawford, 2018). Furthermore, some authors argue that system transparency might (1) inhibit user privacy and (2) enable gaming of the system (Ananny & Crawford, 2018; Veale et al., 2018).

Further work in CS literature aims to establish guidelines about which parts of intelligent systems should be transparent and how the corresponding features should be designed. Diakopoulos (2016) names the following categories that might be disclosed: (1) human involvement (e.g., purpose of deploying system, responsibility inside organization), (2) utilized data (e.g., quality measures such as completeness and timeliness, conducted pre-processing steps, accessibility, privacy implications), (3) model characteristics (e.g., statistical assumptions, input features, weights), (4) inferencing (e.g., performance measures, confidence in predictions). Regarding the design of transparency features, current research in computer science mainly aims to develop explanation facilities for statistical ML algorithms. Transparency features typically serve one of two purposes: (1) visualizing the inner workings of learned ML models (e.g., intermediate representations), and/or (2) providing intuitions about the rationale behind individual predictions (local interpretability) or model workings (global interpretability) (Guidotti et al., 2019). Recent studies concerning model visualization mainly aim to develop transparency facilities for deep neural networks which have achieved state-of-art performance on tasks such as image recognition or language modeling and are naturally hard to understand for humans because of their inherent complexity (often containing millions of learnable parameters). Common approaches intend to visualize the learned representations in hidden layers of neural networks, e.g., they provide intuitions about which specific objects are detected by kernels in a convolutional neural network (Carter et al., 2019). In natural language processing, similar techniques are applied to visualize how recurrent neural networks memorize characteristics of long input sequences for making predictions, e.g., about the next word in a sentence (Madsen, 2019). Regarding the explanation of single predictions and model workings, most approaches use attribution techniques, which allow to measure the relative importance of input features as well as features detected in intermediate representations with regard to the model output (Olah et al., 2018). Techniques like LIME and SHAP explain predictions by quantifying the influence of each input feature and are well established in practice, especially for models working on tabular data (Lundberg & Lee, 2017; Ribeiro et al., 2016). Both techniques work for any supervised ML algorithm and are thus called modelagnostic approaches. Early stage research aims to provide similar mechanisms for other data types and more complex neural networks, e.g., by relating human-understandable concepts to features learned in opaque systems and quantifying their importance for predictions (Kim et al., 2018).

3.2.3 Information Systems Perspective on Transparency

Transparency of intelligent systems has also been a theme of information systems (IS) literature. Early work on transparency mechanisms dealt with Knowledge-Based Systems (KBS). Although these systems oftentimes do not rely on statistical ML methods and thus do not contain self-learning mechanisms, we include prior studies from this research area as some theoretical insights derived from transparency research in KBS should be relatable to modern ML-based systems. In KBS, transparency is mainly achieved via different types of explanations, e.g., rule traces or justifications (Ye & Johnson, 1995). Implementation of these explanation facilities is based on established decision-making theory, e.g., Cognitive Effort Perspective and Toulmin's Model of Argumentation (Dhaliwal & Benbasat, 1996; Gregor & Benbasat, 1999). Furthermore, Dhaliwal and Benbasat (1996) develop a theoretical framework for empirical evaluation of user interaction with explanation facilities, incorporating user perceptions, learning effects and performance measures for judgmental decision-making. Subsequent studies rely on this framework and show that use of explanations improves performance on specific tasks in a cooperative problem solving setting (Gregor, 2001). Moreover, differences between requirements of expert and novice users are established. Whereas novice users more often use explanation facilities for learning, expert users mainly employ these mechanisms in order to verify conclusions (Mao & Benbasat, 2000). More recent IS research examines explanations in recommender systems which more often utilize modern statistical ML models. For recommender systems, integration of explanation facilities is shown to increase decision efficiency and effectiveness, perceived transparency and user satisfaction (Gedikli et al., 2014). The authors also examine differences between multiple transparency mechanisms, varying degree of personalization and integration of content data. They find that non-personalized content-based explanations most positively influence decision effectiveness, whereas more efficient interfaces (e.g., average ratings) result in increased decision efficiency. Another study shows that perceived transparency of recommendation agents positively influences perceived usefulness and enjoyment, mainly through positive effects on affect- and cognition-based dimensions of trust (W. Wang et al., 2016). A positive relationship between perceived transparency and trusting beliefs is confirmed by Wang and Benbasat (2016). They find that perceived transparency positively influences trust in competence, benevolence and integrity of recommendation agents. Although these findings are valuable for our research context, transparency research on intelligent systems is rare so far. Recently, Sidorova and Rafiee (2019) name lack of algorithmic transparency as an inhibitor of AI adoption. They explicitly list undefined data sources and the opaque, complex nature of modern algorithms as problematic aspects regarding this specific technology. Only a few studies explicitly address how transparency features for intelligent systems should be designed (Chai & Li, 2019; Fernandez & Provost, 2019; Martens & Provost, 2014).

3.3 Theoretical Framework

Our study can be divided into two parts, as we aim to find evidence for (1) consumer's overall WTP for transparency features of intelligent systems (RQ1) and the underlying mechanism to establish their purchase intention (RQ2). Therefore, in the first subsection we discuss why we expect consumers to exhibit WTP for transparency features for intelligent systems. The second subsection then introduces our research model for explaining the mechanisms behind the purchase decision.

3.3.1 Willingness to Pay for Transparency Features

As mentioned in the introduction, consumers are oftentimes left in the dark regarding how predictions about personal characteristics (e.g., creditworthiness) are made by intelligent systems. In most cases, the utilized data and details about the applied algorithms are not disclosed to the consumer, thus creating an information gap between the consumer and the provider of the intelligent system (Burrell, 2015; Martens & Provost, 2014; T. Miller, 2019). We argue that transparency features would constitute an added value for consumers because they provide a way to close this information gap. In detail, transparency features could include information about the intelligent system provider (e.g., certifications), utilized data (e.g., sources, pre-processing steps), model characteristics (e.g., applied algorithm, input features, feature weights) and inferencing (e.g., performance, confidence in predictions) (Diakopoulos, 2016). Having established the interest consumers should have in transparency features, the question remains why they would also be willing to pay for them. Here, we build on findings from privacy research, where examining how consumers value their personal data has been a theme for many years. Multiple studies show that consumers put meaningful monetary value on their personal information (Grossklags & Acquisti, 2007; Krasnova et al., 2009; Schreiner & Hess, 2015; Wagner et al., 2018). To the best of our knowledge there are no studies examining WTP in the context of intelligent systems. However, as intelligent systems frequently process personal information and their predictions are oftentimes related to personal characteristics, we assume that consumers will exemplify comparable behavior as in privacy contexts. Based on these observations, we expect that consumers are willing to purchase transparency features for an intelligent system that judges their personal characteristics.

3.3.2 Mechanisms behind Purchase Decision

Our study aims to explain whether and why consumers are willing to pay for transparency features in a finance context. In order to account for consumer's actual intention to purchase these features, we draw inspiration from Theory of Planned Behavior (TPB). TPB postulates that the behavior of individuals is induced by behavioral intentions, provided that the intention itself results from the individual's attitude towards the behavior as well as from subjective norms and perceived behavioral control regarding the according behavior (Ajzen, 1991). We mainly use TPB as a starting point because this theoretical framework can be flexibly adapted and offers the possibility to add new variables (Venkatesh et al., 2003). With the goal of keeping our theoretical model concise we focus on relevant constructs from the TPB framework. To the best of our knowledge, this study is the first relating TPB to consumer's intention to purchase transparency features in intelligent systems. However, TPB has previously been used in WTP and purchasing contexts, making it suitable for our study as well (George, 2004; Hansen et al., 2004; Schreiner & Hess, 2015). Because of our focus on explaining purchase behavior, traditional theoretical models with a focus on adoption, e.g., the Technology Acceptance Model (TAM) or Unified Theory of Acceptance and Use of Technology (UTAUT), are not suitable for this study (Davis et al., 1989; Venkatesh et al., 2003).

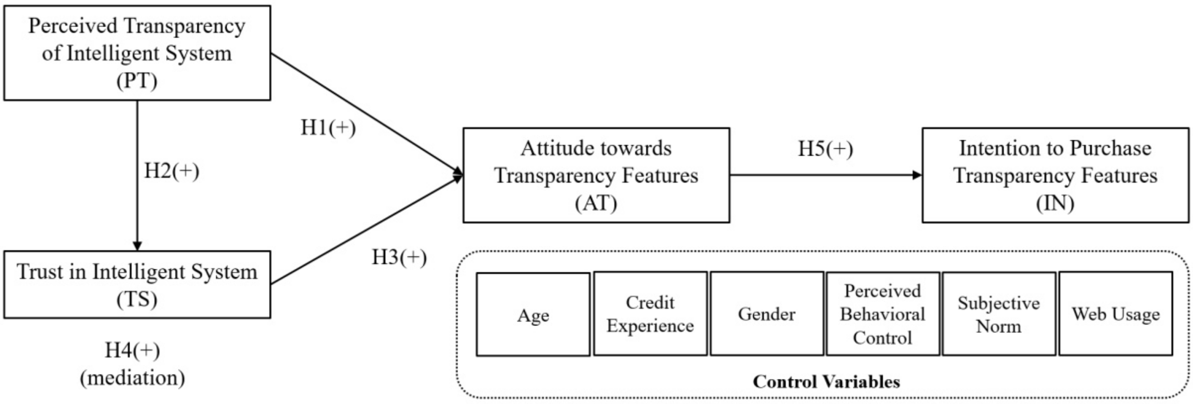
Perceived Transparency of Intelligent System (PT)
H1(+)
H2(+)
Attitude towards Transparency Features (AT)
H5(+)
Intention to Purchase Transparency Features (IN)
Trust in Intelligent System (TS)
H3(+)
H4(+) (mediation)
Age
Credit Experience
Gender
Perceived Behavioral Control
Subjective Norm
Web Usage
Control Variables

Figure 2: Research model for mechanisms behind purchase decision

Figure 2 shows our research model, including all relevant model constructs and control variables. In accordance with previous work on transparency in similar contexts, we define perceived transparency of intelligent system (PT) as the consumer's capability to understand the inner workings of an intelligent system, including assumptions and characteristics that determine its outputs (W. Wang & Benbasat, 2016). Moreover, we examine influences of trust in the intelligent system (TS) onto attitude towards transparency features (AT), which we define as the degree to which a consumer favorably or unfavorably assesses transparency features. Also adopted from TPB, we use intention to purchase transparency features (IN) as our main predicted variable. Finally, we include the two remaining TPB constructs as control variables: subjective norm (SN) resembles perceived social pressure to purchase transparency features and perceived behavioral control (PBC) is defined as perceived ease or difficulty of purchasing transparency features.

In order to examine the mechanisms that underlie this behavior, we rely on TPB and establish antecedents for consumer's attitude towards the transparency features. Increased perceived transparency in the form of additional insights about inner workings of a system constitute the main benefit for consumers when evaluating whether to buy transparency feature packages. Hence, we hypothesize that perceived transparency will have a positive influence on consumer's attitude toward transparency features. Furthermore, we expect this effect to be mediated by trust into the intelligent system as a whole. The more information consumers receive about how a system works, the more they trust outputs of the respective system. This assumption can be justified by previous studies in transparency research for related contexts (W. Wang et al., 2016; W. Wang & Benbasat, 2016), and comparable work regarding the evaluation of privacy-enhancing features (Schreiner & Hess, 2015). An increase in trust into the intelligent system should then lead to a more positive evaluation of the transparency features, as assessed by studies examining different contexts (George, 2004; Schreiner & Hess, 2015; Suh & Han, 2002; Wu & Liu, 2007). All in all, we postulate the following hypotheses:

H1. PT will positively affect consumer's AT.

H2. PT will positively affect consumer's TS.

H3. TS will positively affect consumer's AT.

H4. PT will positively affect consumer's AT because of increased TS.

Following the TPB framework, we expect that a positive attitude towards transparency features constitutes the main predictor for their intention to purchase these features. Previous studies that apply TPB to the related context of internet purchases confirm this assumption (George, 2004; Hansen et al., 2004). Furthermore, we control for subjective norm and perceived behavioral control, which are additional predictors for consumer's intentions according to the theory (Ajzen, 1991). Therefore, we assume that:

H5. AT will positively affect consumer's IN.

3.4 Research Design and Data Collection

3.4.1 Study A: Measuring Willingness to Pay

Following the recommendations of Karahanna et al. (2018), we conducted an online lab experiment since transparency of intelligent systems has not been monetarized yet and new insights into the evaluation of this technology feature are still necessary. By using this method, we ensured the highest degree of internal validity possible as the boundary conditions of an experiment can be determined precisely (Karahanna et al., 2018). In the beginning, the respondent was briefed that she/he needs to apply for credit of 50,000 € (corresponds to mean credit amount in Germany). Furthermore, respondents were told that their financial institution employs an intelligent system, i.e., AI which collects and evaluates online data from sources like corporate websites, civil registers and credit agencies. This approach is in line with real rating providers such as Branch, Kreditech or FICO, that use ML algorithms to specify credit scores (Branch, 2019; FICO, 2019b; Kreditech, 2018). All respondents received an initial credit score of 65 points (scale 1-100, 100 = highest creditworthiness). This score was based on a pre-study and ensured that participants were not certain about the final decision of their financial institution, i.e., credit commitment or rejection (see Table 3).

Once the initial score had been provided, participants were offered to purchase additional transparency features. In order to create a realistic setting, the participants received a mock-up containing an example output of the transparency features. The features were designed according to established practices from CS literature (Diakopoulos, 2016; Lundberg & Lee, 2017; Ribeiro et al., 2016) and findings from social sciences about human perception of explanations (Gedikli et al., 2014; Gregor & Benbasat, 1999; T. Miller, 2019). In detail, the transparency features contained information about collected features based on online data from various sources (e.g., job status based on data from corporate website and professional networks) as well as the relative importance of each feature for the final prediction on creditworthiness (e.g., 29%). All participants received a fictional budget of 50 €, which was determined based on a second pre-study (see Table 3). We simulated a real-world purchase decision, i.e., respondents had to evaluate costs compared to the personal value they attach to the transparency package. In order to precisely approximate consumer's real WTP, we employed the Name-Your-Own-Price (NYOP) method with multiple bidding rounds, which is an established method for measuring WTP (Breidert et al., 2006) and has been applied in various contexts (e.g., Chernev, 2003; Hann & Terwiesch, 2003; Spann et al., 2004). Therefore, the applicability in our credit scoring setting could be assumed. According to the method, respondents had to bid for the additional package against an unknown, but fixed threshold price of 15€ (see Table 3) by submitting bids between 0€ and 50€. Moreover, respondents of the study were informed about a total of three bidding rounds that were interrupted as soon as the threshold was surpassed (experiment process is summarized in Figure 3). By including a repeated bidding option, participants were able to adjust to the situation and feel more comfortable as they could include feedback in their bidding behavior and thus overcome the lack of experience concerning the monetization of transparency features (Liu et al., 2016). As NYOP is not an incentive-compatible method on its own, we created an incentive for the respondents by promising them a variable pay-out based on their choices during the experiment (Breidert et al., 2006).

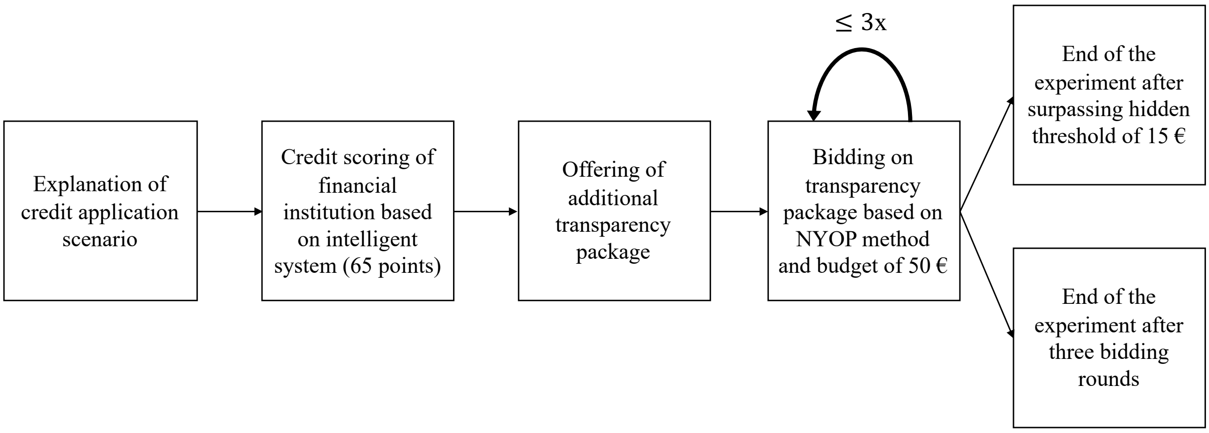
Explanation of credit application scenario
Credit scoring of financial institution based on intelligent system (65 points)
Offering of additional transparency package
≤ 3x
Bidding on transparency package based on NYOP method and budget of 50 €
End of the experiment after surpassing hidden threshold of 15 €
End of the experiment after three bidding rounds

Figure 3: Experiment procedure

In order to set all experiment parameters in the first place, two pre-studies were performed (see Table 3). The first pre-study was conducted to determine a credit score, where it is not obvious whether a credit is granted or not. Therefore, respondents were asked to indicate two values between 0 and 100 points, for which they feel (1) confident that they will be granted a credit and (2) confident that they will be rejected for a credit. This procedure was intended to prevent possible distortions within the experiment that could arise from participant's knowledge about credit acceptance or rejection. In the second prestudy, we employed a direct survey to determine the fictitious budget communicated to the experiment participants as well as the unknown and fixed threshold to bid against. In this regard, we asked respondents of the pre-study to directly report their WTP for a transparency package within a credit scoring scenario. As described above, we assume that the participants of pre-study 2 did not reveal their true WTP (e.g., due to prestige effects). Therefore, the results provide a first guidance to parameter settings but should be critically examined by conducting the experiment and employing the NYOP method.

Table 3: Overview of pre-studies and experiment parameters

	N	Parameter Name	Measure	Parameter Value
Pre-study 1	56	Initial credit score	Median of stated uncertainty range	65 points
Pre-study 2	60	Threshold price	Median of stated WTP	15 €
		Budget	90th percentile of stated WTP	50 €

3.4.2 Study B: Explaining Purchase Decision

After the WTP had been measured by the experiment described above, participants were tasked with completing a survey. Here, we assessed constructs related to our presented research model regarding the purchase decision. Only adapted standard scales were used that are known from and validated in extant TPB literature or that could be derived from the literature concerning system transparency. Suitability of the applied constructs was positively evaluated using a subset of items in the second pre-study conducted with 60 students. In the resulting survey, reflective latent measures were used, operationalized on a seven-point Likert scale: IN, PBC, PT, SN and TS ranged from 1 (strongly disagree) to 7 (strongly agree), while AT was implemented as a semantic differential evaluated from 1 (bad, foolish, unpleasant, dislike) to 7 (good, wise, pleasant, like). Table 4 shows exemplary items of our survey constructs.

Table 4: Examples of construction operationalization

Construct	No of Items	Example of Items	Source
PT	6	With the additional transparency features it gets readily apparent to me how the algorithm generates its prediction.	W. Wang & Benbasat, 2016
TS	5	The intelligent system keeps my best interests in mind.	Koufaris and Hampton-Sosa, 2004
AT	4	Purchasing the additional transparency features is a bad/ good idea.	Taylor and Todd, 1995
IN	3	I intent to purchase the additional transparency features.	Venkatesh et al., 2003
PBC	3	I would be able to purchase the additional transparency features.	Taylor and Todd, 1995
SN	2	People who influence my behavior would think that I should purchase the additional transparency features.	Taylor and Todd, 1995

3.4.3 Sample

The previously described studies were conducted within Germany in March 2019 in cooperation with a market research institute. A total of 223 participants completed the experiment and the survey, resulting in an overall response rate of 76.9%. While selecting the participants, quotas were taken into account in order to reflect the age and gender distribution of consumers using online applications (Eurostat, 2018). The sample consisted of 46.2% female and 53.8% male participants and included a wide range of age groups (18-68 years) resulting in a mean of 38.8 years (SD: 12.6). The majority of respondents were salaried employees (58.5%) and had experience in applying for a credit (64.1%). 79.5% of participants indicated that they use the internet very frequently. Therefore, an appropriate demographic distribution was achieved concerning the setting of both experiment and survey. Hence, our sample allows us to draw inferences concerning our hypotheses for consumers from the general public, at least for the country under study. From this sample, we had to filter 28 participants that failed to pass two included attention checks or showed unengaged behaviour, resulting in a final sample of 195 cases for our analyses.

3.5 Data Analysis and Results

3.5.1 Results of Study A

In order to answer RQ1, i.e., to what extent consumers exhibit WTP for transparency features for intelligent systems, we examined the experiment outcomes. We determined WTP for each respondent by extracting her/his maximum bid out of all three rounds (descriptive statistics are listed in Table 5). We observed a mean WTP value of 21.57€, with a standard deviation of 16.11. Out of 195 respondents, 142 crossed the threshold price of 15€ (72.8%). Thereof, 95 participants surpassed the threshold price in round 1 (48.7%), 38 in round 2 (19.5%) and 9 in round 3 (4.6%). Only 43 participants (22.1%) exhibited bids of 0€ in all three rounds, thus not demonstrating WTP for the transparency feature package.

Table 5: Descriptive statistics regarding measured WTP

	N	Mean	Std. Dev.	25%	Median	75%	Skewness	Kurtosis
WTP	195	21.57	16.11	10.00	20.00	30.00	.31 (.17)	-.80 (.35)

3.5.2 *Results of Study B*

Having established that WTP exists for transparency features, we aimed to examine mechanisms in consumer's purchase decisions using our previously hypothesized research model.

Table 6: Factor loadings and reliability

Construct	Items	Factor Loadings	Cronbach's Alpha
PT	PT1	.797	.942
	PT2	.791	
	PT3	.910	
	PT4	.920	
	PT5	.925	
TS	TS1	.885	.961
	TS2	.774	
	TS3	.809	
	TS4	.853	
	TS5	.911	
AT	AT1	.931	.936
	AT2	.923	
	AT3	.744	
	AT4	.754	
IN	IN1	.966	.971
	IN2	.842	
	IN3	.875	
PBC	PBC1	.931	.809
	PBC2	.631	
	PBC3	.644	
SN	SN1	1.036	.894
	SN2	.558	

Table 7: Factor correlations, reliability and validity measures

	CR	AVE	PT	TS	AT	IN	PBC	SN
PT	.943	.767	**.876**					
TS	.962	.834	.621	**.913**				
AT	.937	.787	.512	.794	**.887**			
IN	.972	.919	.514	.746	.785	**.959**		
PBC	.817	.603	.670	.443	.343	.426	**.777**	
SN	.900	.818	.442	.683	.737	.750	.323	**.905**

Measurement Model. In order to establish that our measurement model was suitable for causal analyses, we performed exploratory factor analysis (EFA) in SPSS 26 and confirmatory factor analysis (CFA) using AMOS 25. During EFA, we dropped item PT6 due to weak factor loadings, all other items were retained. We assessed reliability and validity of constructs by examining factor loadings, Cronbach's alphas, composite reliability (CR) and average variances extracted (AVE) (see Table 6 and Table 7). All model constructs showed highly positive loadings, cross-loadings were smaller than .30 (omitted here for the sake of brevity). Reliability was given, as Cronbach's alphas and CR were above .7 for all constructs, and convergent validity was ensured by AVE greater than .5. Discriminant validity was assured, as square roots of AVE were greater than factor correlations and cross-loadings were smaller than factor loadings. Model fit for the measurement model was also above accepted thresholds, CFI = .98 (> .95), SRMR = .03 (< .08), RMSEA = .05 (< .06). All thresholds are taken from Hu and Bentler (1999).

We tested for common method bias using collinearity diagnostics and Harman's single factor test. Regarding collinearity statistics, we looked at variable inflation factors (VIFs) for each examined relationship in our structural model. No method bias was found, as all VIFs were below the threshold of 3.3 (Hair et al., 2010). In addition, we ran Harman's single factor test and observed that no single factor accounted for more than 50% of total variance. This further indicated that occurrence of common method bias was very unlikely in this study (C. M. Fuller et al., 2016; Podsakoff et al., 2003). Finally, we imputed factor scores for subsequent path analysis using AMOS.

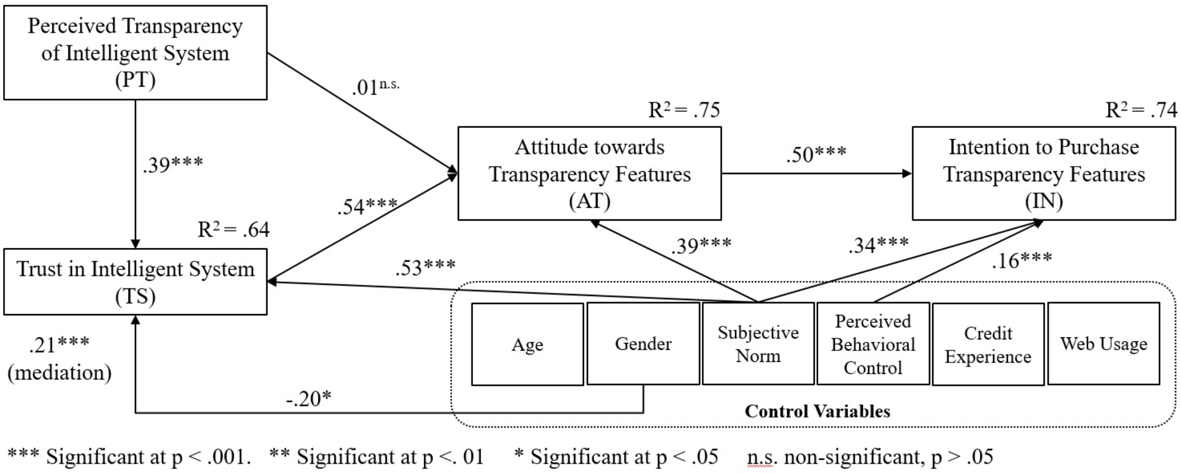
Perceived Transparency of Intelligent System (PT)
.01n.s.
R² = .75
R² = .74
Attitude towards Transparency Features (AT)
.50***
Intention to Purchase Transparency Features (IN)
.39***
.54***
R² = .64
.39***
.34***
.16***
Trust in Intelligent System (TS)
.53***
Age
Gender
Subjective Norm
Perceived Behavioral Control
Credit Experience
Web Usage
.21*** (mediation)
-.20*
Control Variables
*** Significant at p < .001. ** Significant at p <. 01 * Significant at p < .05 n.s. non-significant, p > .05

Figure 4: SEM results

Structural Model. Following our evaluation of the measurement model, we tested our hypotheses by examining path coefficients and their significance (see Figure 4) using AMOS 25. In order to ensure our model's predictive relevance, we again assessed model fit. Model fit was found to be adequate, according to widely accepted thresholds, CFI = .99, SRMR = .05, RMSEA = .06 (Hu & Bentler, 1999). Furthermore, we were able to account for 64 to 75% of the variance in the endogenous constructs TS, AT and IN. Regarding path coefficients, we found no support for H1 as there was no observed direct effect between PT and AT. However, we found that PT significantly influenced TS, with TS significantly affecting AT (supporting H2 and H3). Moreover, H4 is supported, as an indirect effect of PT on AT was confirmed via mediation analysis using bootstrapping (5000 samples), resulting in an statistically significant indirect-only effect over TS, $\beta = .21$, p < .001 (Hayes, 2009; Zhao et al., 2010). We could further establish a significant effect of AT on IN, as proposed by TPB (Ajzen, 1991). Consequently, we can also support H5. Figure 4 also displays statistically significant paths related to our control variables. As expected, PBC and SN positively affected IN (Ajzen, 1991). SN also influenced both TS and AT in a statistically significant way. Furthermore, we found a statistically significant effect of gender on TS. Table 8 sums up our analysis of study B.

Table 8: Overview of tested hypotheses

Hypothesis	Relationship	Support
H1	PT (+) → AT	No, p = .89
H2	PT (+) → TS	Yes, p < .001
H3	TS (+) → AT	Yes, p < .001
H4	PT → TS → AT (Mediation)	Yes, p < .001
H5	AT (+) → IN	Yes, p < .001

3.6 Discussion

This study examined to what extent consumers would be willing to pay for transparency features in the context of intelligent systems (RQ1), and what the mechanisms behind the purchase decision looks like (RQ2). We conducted an online lab experiment with a subsequent survey in order to examine both research questions adequately and increase robustness of results. To the best of our knowledge, we were the first to study WTP for transparency features in intelligent systems. In addition, we developed a research model for the mechanisms regarding the purchase prediction, drawing inspiration from TPB and prior work in related areas like knowledge-based and recommender systems. For the experiment, we developed a real-world scenario in which respondents had to choose whether to purchase an additional feature package that offers insights into algorithmic predictions about their creditworthiness. WTP was measured using the established Name-Your-Own-Price procedure with a total of three bidding rounds and a hidden threshold price of 15€. All experiment parameters for the study were rigorously determined, based on results of two conducted pre-studies. Our experiment unveiled that a large majority of participants exhibited meaningful WTP (median 20€) for the presented transparency feature package. Moreover, our research model for explaining the purchase decision was shown to have high measurement quality and explanation capabilities.

Our study offers significant theoretical contributions regarding research into consumer interactions with intelligent systems. First and foremost, we established that significant WTP among consumers exists for transparency features in intelligent systems through an experiment that simulated actual purchase behavior. Furthermore, we are the first to transfer transparency research to a context different from knowledge-based and recommender systems. Our findings indicate that increasing perceived transparency leads to a positive attitude towards feature packages with this purpose. We showed that an increase in trust towards the intelligent system is the main driver behind this effect, which confirms prior research in the area of recommender systems (W. Wang et al., 2016; W. Wang & Benbasat, 2016). This finding also counters the possible fear that consumers might use the derived insights from transparency features to game the intelligent system. In accordance with TPB, a more positive evaluation of transparency features positively affected the actual intention to purchase these. Additional factors influencing purchase intentions were subjective norms, i.e., social pressure, and perceived behavioral control, in accordance with TPB (Ajzen, 1991). Moreover, transparency could be important to consumers not only in the context of intelligent systems, but for a variety of digital customer experiences. Thus, further transparency research should be conducted in related digital contexts.

Our results also have significant implications for practitioners. First of all, we have shown that WTP for transparency features exists for the real-world scenario of credit scoring. This constitutes a valuable insight for providers of intelligent systems and services when evaluating whether to offer comparable features related to algorithmic transparency. Our findings reveal that inclusion of transparency features might have two main benefits. First, they can be monetized separately from the intelligent service and thus constitute an additional revenue stream in the form of a premium service. Here, results from this study could serve as an entry point for price determination. We found that respondents were willing to spend on average 20€ for such a package which constitutes meaningful WTP. Although these findings are limited to the context of credit scoring, they can still be meaningful anchor points for decisionmakers, e.g., product managers. Second, we showed that increasing perceived transparency has significant positive effects on trust into the intelligent system. This insight can be meaningful to practitioners, because trust has been shown to positively influence important business metrics such as customer retention (e.g., Chiu et al., 2012; Gounaris, 2005; Han & Hyun, 2015). Beyond

WTP, our results point towards the relevance of "fair AI" for consumers. This finding is in line with previous research, as well as company and governmental reports that all name transparency as a key driver for establishing fairness in intelligent systems (Doshi-Velez & Kim, 2017; Google, 2019; ITU, 2018). As algorithmic transparency is seen as particularly important in the public sector (Diakopoulos, 2016), results from this study can also be used for intelligent systems that are employed by public institutions.

Our study is subject to some limitations. First, we focused on examining WTP in the financial context of credit scoring. Although we expect our results to be transferable to related contexts involving intelligent systems (esp., for high-stakes decisions concerning personal characteristics), these contexts might have different outcomes associated with increasing perceived transparency. For example, WTP in these contexts might vary from what we found in this study. Second, we conducted our study within one country. Thus, we are not taking potential cultural differences regarding transparency and WTP into account but would not expect large deviations from our findings. Third, we were not able to establish a direct effect between perceived transparency and attitude towards transparency features. This is probably due to the fully mediated effect of perceived transparency on attitude over trust. Fourth, data for our structural equation model was obtained through a single method of data collection for both independent and dependent variables. Thus, a common method bias could reduce our contributions. Although we conducted statistical tests, namely Harmon's single factor test and collinearity diagnostics, that did not point towards the occurrence of common method bias, we cannot completely rule out its existence.

Based on the results from this study, we see the following opportunities for future research. First, our findings have to be examined in other contexts. Here, we expect transparency of intelligent systems to have similar importance across both B2C and B2B scenarios. In a B2B context we believe algorithmic transparency to have comparable relevance, especially in regulated industries such as banking. Second, further research into transparency approaches for intelligent systems is needed. From a technical perspective, it is important to investigate how complex ML models (e.g., deep neural networks) can be queried for human-understandable explanations. Future research should also draw from previous work in the area of knowledge-based and recommender systems in order to establish best practices for the design of interfaces for transparency features (e.g., Gedikli et al., 2014; Ye & Johnson, 1995). Additionally, consumer behavior regarding transparency of intelligent systems should be observed in relation to other service attributes, e.g., predictive performance. Conjoint analysis could be a suitable research methodology for this, as it still allows to integrate WTP. Privacy research has involved similar studies which future transparency studies could draw inspiration from (Krasnova et al., 2009). Case studies with credit scoring providers, potentially in conjunction with field experiments, would be a further alternative to strengthen external control of our study.

3.7 Conclusion

Advances in AI technology have led to widespread use of intelligent systems for a variety of use cases. Oftentimes, these systems rely on modern ML models, e.g., deep neural networks, making them effectively black boxes. Thus, system characteristics like data sources, input features, statistical models and feature importance for predictions are not revealed to consumers of the system. Consequently, public and private institutions have called for transparency and fairness standards regarding intelligent systems (Google, 2019; ITU, 2018; The Economist, 2018). However, how transparency features can be designed and are perceived by consumers remains largely an open question in both research and practice.

In this study, we investigated whether and why consumers would be willing to pay for transparency features of intelligent systems. Therefore, we conducted an online lab experiment and a subsequent survey with 195 participants in a European country, placed in the context of credit scoring. This allowed us to (1) measure WTP experimentally and (2) develop a research model examining mechanisms of the purchase decision with respect to a real-world scenario. We found that consumers exhibited meaningful WTP (median 20€) for the offered transparency feature package. Furthermore, we observed that perceived transparency of the intelligent system positively influenced the trust consumers have into it, which led to a more positive evaluation of the offered transparency feature package. Our results have significant implications for research and practice. To the best of our knowledge, we are the first to study WTP for transparency features with regard to intelligent systems. Since our theoretical framework shows high explanation capabilities, it can inform further transparency research in various B2B or B2C contexts. Moreover, practitioners can use our results as they offer a new perspective on how intelligent systems providers can monetize their services. On the consumer side, our findings also indicate the relevance of transparency in the context of digital services.

4 Paper B: The Impact of Artificial Intelligence on Individual Performance

Title

The Impact of Artificial Intelligence on Individual Performance: Exploring the Fit between Task, Data, and Technology

Authors

- Timo Sturm, Technical University of Darmstadt, Germany
- Felix Peters, Technical University of Darmstadt, Germany

Publication Outlet

Proceedings of the Twenty-Eighth European Conference on Information Systems (ECIS 2020), A Virtual AIS Conference

Abstract

Artificial intelligence (AI) is increasingly deployed in organizations, allowing information systems (IS) to incorporate self-learning mechanisms. Machine learning (ML) is commonly used as the underlying technology, as it enables IS to derive patters from collected data and perform tasks that were previously reserved for humans. While organizations hope to increase their efficiency and effectivity through adopting AI, the actual linkage between AI use and performance impacts for individuals remains largely overlooked in IS research so far. Therefore, we employ a qualitative research approach to develop a theoretical model for this relationship. In detail, we conduct expert interviews and build on the widely used "task-technology-fit" (TTF) theory. We identify relevant dimensions for the main theory constructs and expand the theory with further components to fit the AI context. Our findings enable future empirical research regarding performance impacts of AI use. Practitioners can use our model to evaluate use cases for AI adoption by considering task, data, and technology characteristics.

Keywords

Artificial Intelligence, Machine Learning, Task Technology Fit, Performance

4.1 Introduction

In recent years, artificial intelligence (AI) has beaten the world's best human Go player (Silver et al., 2017), managed to recognize objects better than the average human (He et al., 2016), and just defeated the world's best professional players in a complex strategic online game (Vinyals et al., 2019). Whereas these examples highlight most advanced technological accomplishments, comparable AI is not only subject to exceptional research projects anymore; AI already influences our lives crucially by helping us to diagnose diseases (Kourou et al., 2015) and control natural disasters (Pourghasemi et al., 2020). Due to its widely recognized transformative potential, organizations have already started to adopt AI in a wide variety of their business functions to increase their efficiency and effectiveness (e.g., Bean, 2019; Forbes Insights, 2018). However, high uncertainty remains on how to manage this new technology to leverage its full disruptive potential (Rai et al., 2019; Rzepka & Berger, 2018). With machine learning (ML) being the major driver of modern AI-based information systems (ISs), the uncertainty of managing AI is further spurred: ML marks an alternative programming paradigm that allows to derive IS functionality from data instead of having humans explicitly translating their solutions into code (Samuel, 1959). AIs that make use of data and ML algorithms – by us referred to as *ML-based AI* – perform intelligent behavior by deriving patterns from data which are then applied to new data to perform actions (Bishop, 2006). The resulting handover of solution design to data-driven algorithms and arising technological particularities (which we will discuss) make it necessary to revisit our existing knowledge on how to manage IS successfully. Especially with AI being frequently praised as a universal panacea for increasing organizations' performance (e.g., Schmelzer, 2019), the actual impact of ML-based AI on organizations' success must be fundamentally questioned and extensively examined.

With today's individuals relying more and more on IS to perform their organizations' tasks, the linkage between ISs and individual performance remains a key concern in IS research (Gebauer et al., 2010; Goodhue & Thompson, 1995). In 1995, Goodhue and Thompson argued that, in conjunction with utilization, information technology (IT) must be a *good fit with the tasks it supports* to positively impact individual performance. They proposed a theoretical model that solidifies this core idea and allows to empirically explore the impact of IS on individual performance (Goodhue & Thompson, 1995). To date, this model is widely known as "task-technology-fit" (TTF) theory. Their results have prompted a dwell of research demonstrating that it is vital for organizations to focus on promoting TTF when managing technology use (e.g., Dennis et al., 2001; Gebauer et al., 2010; Zigurs & Buckland, 1998). Otherwise, organizations may even hinder their individuals' performance, potentially contributing to the organizations' degradation in the long run. In the ML-based AI context, managing this task-technology interplay becomes relevant when individuals place their tasks on AI-produced groundwork: if physicians base their patients' treatments on AIs' medical diagnoses (de Fauw et al., 2018; McKinney et al., 2020) or bankers manage credits based on AIs' credit assignments (Ala'raj & Abbod, 2016; Kruppa et al., 2013), their performance depends on ML-based AIs that augment their work, potentially causing expensive or even deadly consequences if the AIs fail to fit individuals' task requirements. However, can organizations evaluate potential AI-related performance impacts based on traditional TTF constructs given ML-based AI's data-driven design? Or is it required to incorporate resulting ML-based AI particularities (e.g., system transparency or data bias) when deciding on system design to increase individual performance? To the best of our knowledge, it remains unclear to which extent existing knowledge on TTF also applies to ML-based AI or whether new insights are required.

With individual performance being the most fundamental and direct level on which technology's impact on organizations' performance can be explored, it renders suitable to derive a foundation for analyzing ML-based AI's impact on organizations' performance. With this study, we therefore seek to understand the impact of managerial decisions regarding ML-based AI adoption on individual performance. Only recently, researchers have started to investigate the impact of AI diffusion in organizations (Brynjolfsson & Mitchell, 2017; Rzepka & Berger, 2018). Due to this contexts' scarce literature, this study explores factors through a qualitative interview approach with 24 experts that are frequently involved in AI initiatives. Building on TTF as conceptual framework, we aim to answer:

Regarding individual performance, (1) which central characteristics render tasks favorable to be supported with ML-based AI, (2) which central technology characteristics determine ML-based AI use, and (3) which central factors determine the degree of fit between individuals' tasks and ML-based AI?

The remainder of this paper is organized as follows: first, we define ML-based AI, present the TTF theory, and discuss related work. Next, we present our research method, covering our study design and sample. Then, we derive empirical results which we integrate into the TTF theory. To provide a first step towards a theory on the impact of ML-based AI on individual performance, we propose an extended, contextualized theoretical model based on the TTF theory that comprises key characteristics of involved constructs. We conclude by discussing and integrating our key findings into existent research to provide scholars a foundation for future research possibilities and managers essential guidance on how to design ML-based AI initiatives to effectively promote AI diffusion within organizations.

4.2 Theoretical Background

In the following, we first define ML-based AI as a form of AI-based IS and present related work on AI diffusion in organizations. Second, we present the task-technology fit theory and highlight related extensions and applications. Third, we combine both research streams to form our study's objective.

4.2.1 Artificial Intelligence, Intelligent Agents, and Machine Learning

One of the most widely accepted conceptualizations of intelligent behavior in AI research is the one of the "intelligent agent" (David et al., 1998; Legg & Hutter, 2007; Nilsson, 1998; Russell & Norvig, 2016), which is "anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators" (Russell & Norvig, 2016, p. 34). It defines intelligent behavior as an agent function that selects executable actions based on current context information (Russell & Norvig, 2016). This function can be realized with various approaches, e.g., manually defined rules or statistics (Russell & Norvig, 2016). The approach that enabled recent AI advances (e.g., He et al., 2016; Heess et al., 2017; Silver et al., 2017) is ML, the concept of learning from experience through algorithms: ML algorithms are trained on data to create models capturing contained patterns. Trained models are then applied to new data to perform a task (Bishop, 2006; T. Mitchell, 1997). Without ML, solutions are codified entirely by humans, e.g., humans designing rules to define robots' routines (Russell & Norvig, 2016). With ML, solutions result from statistical correlations derived from data, e.g., algorithms that learn how to detect credit card fraud from business transactions (Ala'raj & Abbod, 2016; Kruppa et al., 2013). Thus, ML renders manual programming unnecessary (Samuel, 1959). In our study, we focus on AI as intelligent agents that rely on ML, i.e., *ML-based AI.*

Although calls for more research on collaboration between humans and ML-based AI exist (e.g., Rai et al., 2019), work concerning its impact on individual performance is rare so far. Most existing research approaches the topic from a very practice-oriented perspective, covers aspects such as AI's impact on decision-making (Agrawal et al., 2019), identification of business cases (Fedyk, 2016), or success factors for implementing AI projects (Satell, 2018). Only a few papers contribute to this area on a more abstract and theoretical level. Rzepka and Berger (2018) determine factors that influence user interaction with AI-enabled systems. They mention two fit types that affect human-machine relationships: fit between user and system and fit between technology and task. Brynjolfsson and Mitchell (2017) examine labor implications caused by ML-based AI's diffusion. The authors name criteria for tasks that make them favorable for ML application, e.g., the existence of well-defined inputs and outputs and the acceptance of systems' potential black box behavior. Then, labor implications are discussed by examining effects on established economic factors (e.g., substitution, price elasticity).

4.2.2 Task-Technology Fit

In 1995, Goodhue and Thompson proposed the technology-to-performance chain – today primarily known as TTF theory – as a theoretical model to better understand the linkage between IT and individual performance. They argued that, to positively impact individual performance, IT must *match the tasks well that it supports* when being utilized. They further argue that TTF combined with utilization can thus be applied as appropriate surrogate to predict individual performance (Goodhue & Thompson, 1995). Figure 5 shows the TTF theory which comprises five main constructs: (i) characteristics of tasks that are performed by individuals to turn some inputs into outputs; (ii) characteristics of technologies that support individuals in performing their tasks; (iii) task-technology fit as degree of how well a technology can support an individual's tasks; (iv) utilization as individual's usage behavior of the technology to perform tasks (measurable with, e.g., usage frequency); (v) performance impacts as accomplishment of the individual's tasks with higher performance implying some combination of improved efficiency, effectiveness, and/or quality (Goodhue & Thompson, 1995). According to the theory, task and technology characteristics affect the perceived task-technology fit. This fit then positively impacts performance directly and indirectly via the mediating utilization construct. This theory has been extended and utilized in various contexts. Hereafter, we summarize work related to TTF.

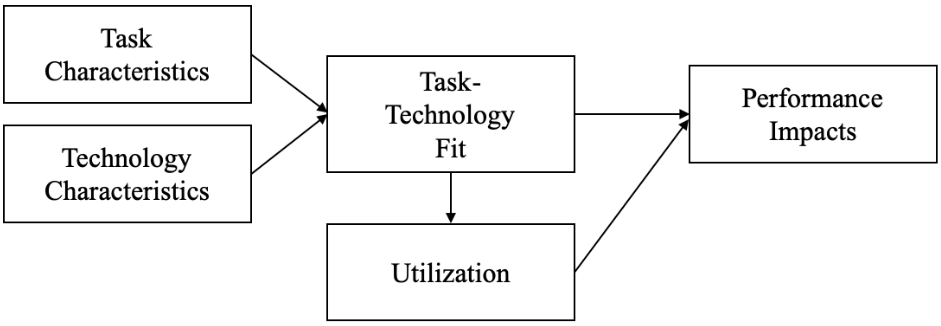

Figure 5: The TTF theory as conceptual base (Goodhue and Thompson, 1995)

Previous TTF research focused on numerous technologies and performance measures. The first technology that TTF was applied to are group support systems (GSSs). Zigurs and Buckland (1998) developed a TTF-based theory to explain GSS effectiveness. In this context, the authors used group performance as target variable and developed different models for five task types: simple, problem, decision, judgment, and fuzzy tasks. For each task, specific GSS functionalities were included as technology characteristics (e.g., communication support, information processing). In later work, TTF was used as one of two factors explaining group performance, the other factor being appropriation effects (Dennis et al., 2001). For this integrated model, TTF was shown to positively influence outcome effectiveness (e.g., decision quality). Further work in the area of GSSs built on TTF to identify an effect of fit between ICT functionality and communication requirements on team performance (Maruping & Agarwal, 2004). This study's target variable was short-term team viability, as measured by satisfaction, team commitment, and group cohesion. In further research, Fuller and Dennis (2009) found that short-term TTF effects on team performance did not sustain in the long term, as poor-fit teams appropriate technology over time, resulting in improved perceived fit and performance. Another context that is widely studied using TTF is the one of mobile IS (Gebauer et al., 2010; Gebauer & Ginsburg, 2009; Junglas et al., 2008; Lee et al., 2007). Here, researchers mainly aimed to develop models based on TTF to examine performance variables such as managerial task performance. To account for the specific particularities of mobile IS, developed models oftentimes included a context construct in addition to the established TTF model components (Gebauer et al., 2010; Gebauer & Ginsburg, 2009). In their research, Gebauer et al. (Gebauer et al., 2010) measured this construct by three variables: degree of distraction, connection quality, and mobility of the user.

TTF has been applied to areas similar to ML-based AI, namely non-ML-based AI, data analytics, and decision support systems (DSS). Here, we refer to non-ML-based AI as AI that is not based on ML but has different underlying technologies (e.g., expert systems). Within this context, previous research aimed to examine individual task performance and intention to use IT. Regarding the first target variable, Wongpinunwatana et al. (2000) developed a model for the impact of fit between an auditing task and an expert system on two variables related to individual performance, namely user's performance on problem solving and user's uncertainty of the correctness of their solutions. Another study integrated TTF and the Technology Acceptance Model (TAM) in order to examine intention to use intelligent agents in web-based auction processes (Chang, 2008). The authors found TTF to be a suitable predecessor to the TAM constructs (e.g., perceived usefulness, perceived ease of use) for the specific tasks of price negotiation and item acquisition. TTF was also used to examine effects regarding use of data in general, and data analytics in particular. It was shown that the TTF model can be used to explain user satisfaction with data (Karimi et al., 2004). Moreover, TTF was established as one of three factors that positively moderate the relationship between data analytics use and firm agility (Ghasemaghaei et al., 2017). Finally, Parkes (2013) developed a model that demonstrated a positive effect of TTF on individual performance in the context of a DSS applied for insolvency legislation.

4.2.3 Summary of Literature Review

The theory of TTF has been used in many different contexts and for a diverse set of technologies. This underlines the suitability of the theory for examining the relationship between technology use and performance impacts on an individual, team, or organizational level. Although TTF has been applied for technologies that have some resemblance to ML-based AI (e.g., expert systems, DSS), findings from these contexts cannot simply be transferred. This is mainly due to two unique characteristics of ML-based AI: First, ML-based AI has to be differentiated from non-ML-based AI approaches, such as expert systems, and other automation technologies as it does not rely on human-defined rules but statistical patterns in data. Second, its focus on providing intelligent behavior rather than aiming for manual extraction of insights distinguishes ML-based AI from approaches like data mining or analytics. Hence, existing research on TTF is not sufficient for the context of ML-based AI. Since there is not enough evidence available regarding ML-specific factors that influence the TTF, we employ an explorative focus for this study. Here, our goal is to identify the most important TTF factors to enable empirical research in this area. In the following section, we will describe the applied methodology.

4.3 Qualitative Research Methodology

With this study, we aim to provide initial evidence regarding general factors affecting the impact of ML-based AI on individual performance mediated by TTF. To achieve this, we questioned experts from operational and managerial levels of different organizations. As justified above, we chose to pursue an explorative approach using interviews to study particularities associated with the use of ML-based AI in this particular context (Flick, 2004). Following Weber (1990), content analysis can be used to evaluate collected qualitative data, making it suitable to assess open-ended questions. We thus apply content analysis by following the steps proposed by Hsieh and Shannon (2005): First, we chose to use the TTF theory as a conceptual basis for our investigations. We made this decision as the TTF theory represents a widely accepted theory which has been empirically proven in many different contexts and focuses on performance impacts of the interplay between tasks and technologies in which we are interested in. We extracted its main constructs as initial categories for potential factors. Second, we conducted and recorded the interviews. Third, we transcribed, coded, and analyzed the interviews considering studies related to ML-based AI's particularities through triangulation (Hsieh and Shannon, 2005), including the rather practical-oriented studies which we presented as related work above. Thus, we combine directed and conventional analysis, where the directed approach aims to draw on codes extracted from existent theory (i.e., the TTF theory) and the conventional analysis aims to derive information directly from gathered data, since we focus initial evidence regarding factors associated with ML-based AI in the context of TTF (Hsieh and Shannon, 2005).

4.3.1 Research Design

We conducted semi-structured interviews with experts of different organizations and varying experience in using ML-based AI within organizational contexts and used these interviews as our key information source. While doing so, we used the principles proposed by Sarker et al. (2013) to guide our interview preparation and execution. Prior to each interview, we discussed our definition of ML-based AI and a set of related example applications with each expert to ensure a shared understanding. During the interviews, we used open questions to enable experts to freely share experiences and views related to our research objective. We designed the interview questions along the TTF theory by varying the questions' focus on the different TTF constructs to explore relevant task, technology, and fit characteristics and related dependencies both in isolation and in combination with one another. In addition, we used the above highlighted TTF and ML-based AI literature to further shape the questions' focus. As a result, our interview guide covers five sections. The first section targets general information about the experts' position, responsibilities, and past experiences with applying ML-based AI in organizational contexts. While this section was primarily designed to familiarize the experts with the interview situation, many statements already provided useful insights as some experts began to mention value and challenges of conducted AI initiatives. The second section focuses on exploring characteristics that are special to ML-based AIs. To achieve this, we primarily ask to describe problems that are suited to be solved with ML-based AI and to differentiate them from manually programmed solutions. The third section aims at organizational requirements as well as organizational and technical challenges related to data and algorithms for creating ML-based AIs. The fourth section focuses on how organizations identify usage scenarios for applying ML-based AIs in their organizational processes. Finally, the fifth section explores achieved and pursued benefits as well as potential risks and negative consequences related to the adoption of ML-based AI in organizational processes. Resulting from the pursued semi-structured approach, initially defined questions were gradually adjusted to meet each expert's individual expertise and to develop the focus during the interview process.

4.3.2 Data Collection and Coding Concept

We based the selection of the experts on a key informant approach. To comply with the rules of data triangulation, we included both provider and user firms (Flick, 2004). We conducted 23 interviews with 24 experts within Europe and Northern America, including fifteen experts from provider and nine experts from user firms (i.e., firms that mainly purchase AI products). One interview included two experts. During the last five interviews, we noticed that additional data discontinued to add new insights which implied that we had reached theoretical saturation (Flick, 2004) and therefore decided to stop interviewing. The interviews were held face-to-face or by telephone and lasted 56 mins on average. They were conducted from December 2018 to April 2019. With our interviews, we aimed to capture experiences related to both technical and organizational topics to avoid an elite bias (Miles et al., 1994) and to enable a combination of both viewpoints which is essential to the TTF theory. All experts work or have worked as data scientist and thus have basic to advanced knowledge in data analysis. Our sample includes data scientists, managers, technical consultants, presales consultants, and developers that are frequently involved in AI initiatives. Each expert regularly deals with the implementation of prototypical or productive systems in different organizational contexts, being especially involved in conducting data exploration and management, algorithmic design and evaluation, and use case identification and definition. The experts' experiences with AI initiatives comprise 19 industries with special focus on the finance (48%), manufacturing (48%), health care (29%), railway (29%), and automotive (24%) industries. Each expert has three to twelve (mean: six) years of experience in one to ten (mean: three) different industries. Table 9 provides detailed information on the involved experts.

Table 9: Experts who participated in the study

ID	Age	Gender	AI Experience	Profession	Firm	ID	Age	Gender	AI Experience	Profession	Firm
i1	51	Male	6 yrs. in 2 ind.	DS	P	i13	32	Male	7.5 yrs. in 1 ind.	DS	U
i2	39	Male	6 yrs. in 2 ind.	DS	P	i14	32	Male	10 yrs. in 4 ind.	TC	P
i3	33	Male	7 yrs. in 5 ind.	M	P	i15	52	Female	4 yrs. in 3 ind.	DS	P
i4	31	Female	7 yrs. in 5 ind.	PC	P	i16	32	Male	10 yrs. in 7 ind.	M	P
i5	44	Male	3 yrs. in 1 ind.	D	U	i17	30	Male	5 yrs. in 6 ind.	PC	P
i6	28	Male	3 yrs. in 6 ind.	PC	P	i18	37	Male	5 yrs. in 3 ind.	DS	P
i7	30	Female	3.5 yrs. in 4 ind.	TC	P	i19	36	Male	8 yrs. in 1 ind.	M	U
i8	33	Male	6 yrs. in 10 ind.	M	P	i20	35	Male	12 yrs. in 2 ind.	DS	P
i9	35	Male	5 yrs. in 3 ind.	TC	P	i21	37	Male	7.5 yrs. in 3 ind.	DS	P
i10	32	Male	9 yrs. in 4 ind.	DS	U	i22	41	Male	3 yrs. in 1 ind.	D	U
i11	25	Female	3 yrs. in 1 ind.	M	U		39	Male	3 yrs. in 1 ind.	D	U
i12	34	Male	6 yrs. in 3 ind.	DS	U	i23	35	Male	6 yrs. in 4 ind.	TC	U

D: developer / DS: data scientist / M: manager / PC: presales consultant / TC: technical consultant / P: provider firm / U: user firm

All interviews were recorded and transcribed in agreement with the interviewees. Following the recommendations in Saldaña (2009), we conducted two coding cycles using the NVivo 12 software to evaluate the transcripts. In the first cycle, we employed attribute coding, hypothesis coding, and descriptive coding. Attribute coding was used to extract information about the collected data, e.g., participant and organization characteristics. Subsequently, hypothesis coding was conducted with the aim of identifying relevant dimensions for the original TTF constructs, i.e., assigning codes to task, technology, and task-technology-fit characteristics. The first coding cycle was concluded by applying descriptive coding to identify additional constructs and construct dimensions that might extend the base theory (here, TTF). The second coding cycle consisted of pattern coding, which was used to condense the identified codes into a smaller number of mutually exclusive and collectively exhaustive constructs and dimensions. The coding process was validated in discussion between five IS researchers and student assistants. Furthermore, we incorporated additional data sources, i.e., articles on ML-based AI use (Fedyk, 2016; Brynjolfsson and Mitchell, 2017; Satell, 2018; Agrawal et al., 2019), to compare our findings with existent knowledge (see section 4), eliminating any ungrounded discrepancies. Thus, research rigor was ensured by performing both data and investigator triangulation (Flick, 2004).

4.4 Results

With our study, we found first evidence for key characteristics of tasks, data, ML-based AIs, and related fits that likely affect utilization and individual performance. We solidified our findings based on the TTF theory and propose an extended, contextualized theoretical model which is illustrated in Figure 6.

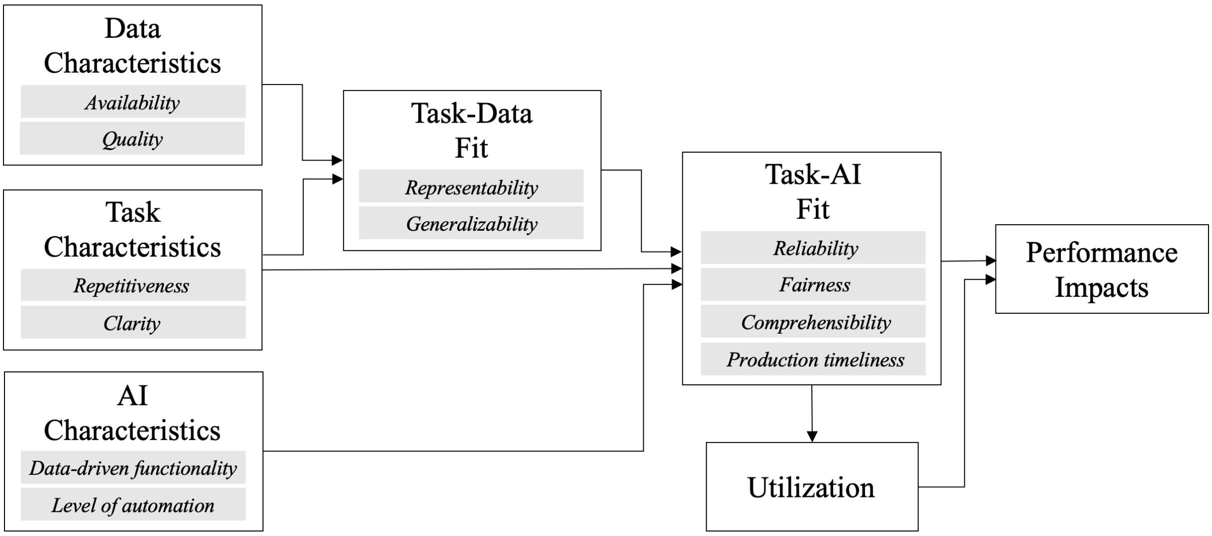

Figure 6: Extended and contextualized theoretical model of TTF in the context of ML-based AI

As utilization can be well measured in empirical studies with actual AI users (e.g., through actual usage frequency), we chose to not investigate utilization because experts' assessments of utilization may be of less value. Instead, we focused on exploring the impact on individual performance that results directly from the fit between tasks and ML-based AIs which is rather difficult to measure. However, following the original TTF theory, it is likely that an impact of task-AI-fit on utilization exists. We thus leave it to future studies to explore the impact on utilization in more detail while we abstract this effect hereinafter. Due to ML-based AIs' strong dependence on data, the experts clearly stressed the importance of the availability of high-quality data for implementing ML-based AIs. As ML-based AIs that support individuals' tasks must act on data collected through or related to the tasks' executions, the experts frequently highlighted that organizations must understand how well their data can actually describe their tasks before they plan to support them with ML-based AIs. Only as a next step, it makes sense to assess whether an ML-based AI of sufficient quality can be derived from organizations' data. We therefore added data characteristics as an additional construct to reflect different data properties' impact on the interplay between organizations' data and the tasks it aims to describe. Moreover, to characterize this interplay and to include related effects, we introduced task-data fit as a further construct. Throughout our interviews, task and technology characteristics emerged that appeared to be central to the use of ML-based AIs. We therefore contextualized the task and technology characteristic constructs to hold such related characteristics. To indicate the specialized focus, we refined "technology characteristics" with "AI characteristics". Due to the mentioned importance of data, we found taskdata fit to impact task-AI fit besides task and AI characteristics. Lastly, we contextualized "tasktechnology fit" as "task-AI fit" and assigned characteristics that emerged to mainly determine task-AI fit and its impact on individual performance. Below, we will discuss each construct in more detail.

4.4.1 Data Characteristics

Availability. Data availability was viewed as a major concern in nearly all interviews and literature as it limits the useable data basis to describe task executions (Fedyk, 2016; Satell, 2018; Agrawal et al., 2019). According to the experts, organizations face two major issues that comprise availability. First, organizations must understand which data they already capture and which they could further collect:

> *"ML-based AI is hungry for data. If you're planning something like this, you need to think about where I'm staying regarding data collection and digitalizing my processes. Even if it's a paper that moves around or somebody clicking somewhere – is there a digital system that captures it in form of data?" (i9)*

Second, organizations must comprehend which captured and capturable data is actually accessible as data ownership of internal and external parties emerged as major obstacle for data access. Especially data privacy restrictions (e.g., of EU's GDPR) often render individuals' data inaccessible if it captures sensitive information. If organizations hold data owned by other organizations, its access is likely legally restricted to clearly defined purposes while organizations' internal parties (e.g., teams or departments) may further restrict the use of data managed by them:

> *"How sensitive is the data? Often, we simply didn't get the data. All our concepts were great, but in the end, we could not use the data due to privacy restrictions." (i5)*

> *"When they ask the other team, they say: 'No, we have our own system. Don't touch that!'. So, this data cannot be accessed." (i12)*

Quality. Both the reviewed literature and our experts frequently highlight that even if organizations hold much available data related to task executions, its quality determines its actual informativeness (Clarke, 2016; Ghasemaghaei et al., 2018; Agrawal et al., 2019). However, organizations' data is often incorrect, imprecise, incomplete, or hard to combine, which reduces the truthfulness and coverage of captured information. Furthermore, organizations' data is often stored in different forms, granularity, and split across multiple sources which often leads to coarse and non-combinable data, potentially reducing the extent to which organizations' data can capture elements involved in task executions:

> *"As of data quality, you basically want to ensure that each data point captures something that actually happened, there are no duplicates, no missing data, and the data is truthful and doesn't get mixed up somehow due to processing errors. If it is encrypted or compressed, it can result in some loss of information." (i10)*

Hence, we posit:

Proposition 1: *In the context of ML-based AI, data availability and quality likely are the central data characteristics that impact task-data fit.*

4.4.2 Task Characteristics

Repetitiveness. We found that ML-based AIs are generally used to support individuals' tasks by doing task-related groundwork in an automated manner, i.e., by carrying out subtasks to provide interim results of individuals' tasks (Brynjolfsson and Mitchell, 2017; Traumer et al., 2017; Agrawal et al., 2019). This allows individuals to base subsequent subtasks on the AI's output to complete their tasks. Literature and experts agree on ML-based AI being a tool for automation used by organizations to reduce workload in their individuals' tasks (Brynjolfsson and Mitchell, 2017; Satell, 2018). Thus, we found that the more repetitive supported tasks are, the greater AIs' potential impact on individuals' workload may become. High-level repetitiveness therefore likely amplifies the effect on individuals' performance that results from the actual fit between individuals' tasks and supporting ML-based AIs:

> *"We assess the task's frequency: Where does an expert lose a lot of his time due to a repetitive task? That's where we have a big automation potential for which I may use ML. It has less impact in very diverse, very versatile, very specialized task contexts." (i18)*

> *"High value is where AIs can take care of a lot of repeated things most of the time, so that you only need to address the last 20% of situations that are somewhat difficult [for the AI]." (i8)*

Clarity. According to our experts and reviewed literature, ML-based AIs are most suitable to support tasks which comprise some uncertainty on how to transform given input into potential output, i.e., non-trivial tasks in which the actual connection between input and output remains largely unclear (Brynjolfsson and Mitchell, 2017; Agrawal et al., 2019). It further became apparent that this uncertainty generally results when tasks allow a great number of potential input-output connections from which the most optimal one must be chosen. The experts view them as non-trivial, as comparing all possibilities is at least very tedious or even impossible while the best option remains non-obvious. It got apparent that individuals more strongly rely on gathered experience and instinct when executing such higher complexity tasks. The experts further agree that using ML-based AI to support tasks for which humans can articulate a sufficient solution by defining a clear set of rules may lead to a worse fit with the supported task as ML-based AIs introduce characteristics resulting from giving up control over systems' operating principles (see TTF characteristics). The experts even view it as second-choice tool if it is possible to create a rule-based IS that produces comparable results to retain better system control:

> *"It should be problems where the functional relationship is widely unknown, so that I cannot program it directly. They must be so complex that one cannot recognize a correct solution without more ado. [...] If possible, I would always prefer to use the known rules because then I know that the things will happen that I would like to see and do not have to hope that the algorithm learns what it should learn instead." (i1)*

Therefore, we propose:

Proposition 2: *In the context of ML-based AI, repetitiveness and clarity likely are the central task characteristics that impact task-data and task-AI fit.*

4.4.3 Task-Data Fit

Representativity. In our interviews and literature review, it became clear that organizations' available data must be as representative as possible for some task's execution, i.e., reflect as much relevant aspects as possible that determine a task's real-world execution (Traumer et al., 2017). If it misses or falsifies relevant aspects or describes aspects that are irrelevant for the task's execution, contained correlations may miss to reflect or may even imply wrong relations in the task's execution. It therefore became apparent that the representativity of data does not only depend on organizations' capability of collecting data but also on the nature of the task. Especially, if individuals use general knowledge or subjective judgment to make decisions involved in a task, related data likely insufficiently represents the task's execution when the task itself does not allow to collect data that grasps such elements. Thus, the experts highlight that a lack of representativity of organizations' available data for tasks' execution may mislead resultant ML-based AIs in executing subtasks to support individuals:

> *"In machine learning, the data you have is really at the core of the problem of how to define it and, more importantly, how to solve it. I think any solution can only be as good as how representative the data is of the problem that you're trying to solve. If you're trying to build an AI for predicting customer churn, but you don't have any data about customer complaints, then you might not be so successful." (i10)*

> *"In the best case, an AI may extract many or all possible information from the data that describes a task execution. However, this also depends on the nature of the task. For example, if you do not have the right sensors, the AI may not be able to derive important information. If there are important occurrences that are not represented in the data, then the AI will have no chance to determine these based on the data." (i2)*

Generalizability. Both literature and experts frequently state that even if organizations are able to capture tasks with representative data, the task's nature itself may render it non-generalizable (Fedyk, 2016; Brynjolfsson and Mitchell, 2017; Satell, 2018). This is the case, if aspects related to task execution change significantly over time. This temporal change may render available historic data insufficient to describe today's task execution when derivable relations do not hold true anymore and thus cannot be used to generalize task execution:

> *"But in the AI world, there is the extra level of complexity: the data is always funny and you never know whether or not the distributions of data are going to change over time or if the AI problem itself is going to change over time both from my data but also from my business point of view." (i8)*

Besides, if involved decisions are rather driven by individuals' instincts instead of knowledge or gathered experience, their task execution may follow no consistent logic, rendering the task nongeneralizable due to the lack of derivable reoccurring structures of the task's execution:

> *"Of course, you have to expect a fitting pattern in the data. If you expect no connection to be existent at all, if everything is random, you cannot hope that ML will find any patterns. So, you have to expect that patterns exist, and they have to be so complicated that you cannot manually recognize them easily." (i1)*

We thus posit:

Proposition 3: *In the context of ML-based AI, representativity and generalizability likely are the central task-data fit characteristics that impact task-AI fit.*

4.4.4 AI Characteristics

Data-driven functionality. The functionality of ML-based AIs bases on derived data patterns instead of having humans manually specifying a rule set that defines the IS's functionality (Samuel, 1959; Bishop, 2006). The experts highlight that this alternative programming approach changes the possible customization of the resulting IS's system behavior. While rule-based ISs allow to manually adapt their system behavior by adding, modifying, or removing rules, ML-based AIs' behavior can only be manually adapted by adding human-defined rules that act on the AI's input and output. Both experts and literature stress that most ML algorithms do not allow a manual adaptation of the core of an ML-based AI's behavior, i.e., its pattern-based agent function that connects the AI's inputs and outputs (Brynjolfsson and Mitchell, 2017). Instead, a new AI has to be created that bases on other data, algorithm, or parameters of the algorithm to modify its behavior. Therefore, the customization can only be performed indirectly by organizations. However, this likely changes the AI's overall operating principles instead of adapting targeted functionality in isolation:

> *"If you have certain cases that the AI treats in a wrong way, then it may be an incredible effort to change the AI's behavior in such a way that it treats them correctly without changing its treatment of other cases too. Without ML, I could simply add some if-else rule to adapt the system's behavior. With ML, I cannot simply treat certain cases in isolation, but actually have to solve the entire problem from the beginning again." (i17)*

Due to this, the experts further state that an ML-based AI's functionality gets shaped by the characteristics of the ML algorithm utilized to create it. This includes the algorithm's transparency, capturable complexity, capabilities of handling data bias, and latency that appeared to form the ML-based AI's reliability, fairness, comprehensibility, and production timeliness (i.e., task-AI fit characteristics).

Level of automation. ML-based AIs support individuals by automating parts of their tasks. We found that this support can be realized in different forms depending on the ML-based AI's level of automation (Traumer et al., 2017; Agrawal et al., 2019). Throughout our interviews, the experts frequently discussed two main forms. As of a rather low level of automation, ML-based AIs may support individuals by offering a list of recommendations ordered along the AI's estimated likeliness of being an accurate output for the subtask. The number of included recommendations varies with the minimum of offering a single recommendation. At a rather high level of automation, ML-based AIs may also support individuals by autonomously acting upon their own derived subtask-related output. With a higher level of automation, the individuals appear to become more dependent on ML-based AIs. As little automated ML-based AIs allow individuals to explore their recommendations before basing their entire tasks on the AI's output, highly automated ML-based AIs rather force individuals to exploit the AI's output for their resulting task execution. Thus, with little automated ML-based AI, individuals have a better chance to evaluate the fit between the ML-based AIs' output and the individuals' task execution (e.g., evaluate the ML-based AIs' output correctness). One expert exemplified this as follows:

> *"In the context of predictive maintenance: If my AI has identified a failure, it may say: 'In the next two weeks, your pump will be leaking. Do something!'. Instead, it could also recommend: 'Someone has to go there.' or it could even send someone directly." (i3)*

Thereby, we propose:

Proposition 4: *In the context of ML-based AI, data-driven functionality and level of automation likely are the central AI characteristics that impact task-AI fit.*

4.4.5 Task-AI Fit

Reliability. As ML-based AIs act on generalized patterns that cannot handle every possibility, they will certainly produce errors at some point (e.g., Bishop, 2006; Brynjolfsson and Mitchell, 2017). Therefore, when organizations consider to supporting tasks with ML-based AIs, they must understand which consequences of potential errors may arise for individuals as they likely perform tasks wrongly if they base them on AIs' erroneous behavior (Agrawal et al., 2019). The experts stress that, to evaluate fit, organizations must therefore understand consequences of AIs' error rate in the task context:

> *"Can I make one error out of hundreds? Sounds very reasonable, but it depends. If I am predicting a cancer patient, I cannot make a false prediction. But if I'm trying to predict whether a customer is going to convert, nobody is going to lose his life. So, there you can actually make more than 20% error." (i12)*

A high AI error rate may reduce the quality of individuals' task outcomes if the errors transfer to the individuals and thus may negatively affect their effectiveness. Individuals' efficiency may also be reduced when they must evaluate the correctness of AIs' outputs and adjust errors, creating additional effort. Besides different error rates, the experts stress that different error types may impact individual performance differently. Organizations should therefore assess the different error types' consequences to understand which error types are more severe in the task context. As ML-based AIs can be designed to favor different types of errors while sacrificing others, the experts highlight that their designed balance of error types should be considered to match the task best, as exemplified in the following quote:

> *"The classic example is the AIDS test. Of course, you'd much rather have a false positive than a false negative, and then you'd say, 'I do it in such a way that I weight a mistake in one direction a hundred thousand times more relevant than the other.' And that's just how it is in the business case. It always depends on the consequences of my decision and you have to balance them in such a way that you achieve the result that you want. Of course, you cannot judge every wrong decision equally. This usually makes no sense from a business point of view." (i20)*

Therefore, the experts suggest that organizations should compare potentially ML-based AI-related saved efforts with possible additional efforts resulting from ML-based AIs' erroneous behavior in the task context, e.g., by measuring the variance of error rates with and without ML-based AI:

> *"If my alternative solution, that was based on humans, had 60% of success and the AI solution is 95%, then it is better than my alternative solution and I'm definitely going for that. [...] You basically compare it to the baseline that you have to identify whether it is the right solution or not." (i9)*

Fairness. If ethnical or social groups are underrepresented or human preferences and prejudices are captured in data, an ML-based AI that is trained on it may discriminate against certain entities due to contained data bias (e.g., Angwin et al., 2016; Chouldechova, 2017). Therefore, experts frequently highlight that organizations must assess whether ML-based AIs may promote discrimination in supported tasks and have to understand which injustices likely result in specific task contexts, as demonstrated in the following quote:

> *"Minorities always come of badly or are not considered at all in an ML model as they are statistically less relevant. That is a big problem and you have to be aware of it to weight minorities correctly in these algorithms. For example, a Portuguese minority in some country may be much more affine for loans which also always reliably repays, but by being a minority, they are less well rated [by the AI]. This means that they will get a bad credit, even though they are actually very good credit customers." (i14)*

However, the experts further emphasize that ML-based AI may also remove existing individual injustice when standardizing the execution of some existing task:

> *"It may be the case that certain data reflects prejudices. I also find it interesting that you can use AI to show which prejudices you had to deal with so far." (i4)*

Thus, organizations must reflect on possible injustice involved in their individuals' tasks to evaluate whether an ML-based AI likely improves or harms its individuals' fairness. As with erroneous behavior, a misfit between the ML-based AIs' fairness and the task context may negatively impact individuals' effectiveness and efficiency when task outcomes reflect injustice and thus require individuals to actively assess and restore justice, while a good fit may positively impact an organization's fairness.

Comprehensibility. Depending on the ML algorithms used to create AIs, their working principles may remain unknown to their users as it is the case with, e.g., neural networks, and thus constitute "black box" behavior (Brynjolfsson and Mitchell, 2017; Guidotti et al., 2018; Miller, 2019). Individuals therefore can have difficulty in assessing ML-based AIs' output and behavior. The experts frequently stress that organizations must therefore understand which degree of comprehensibility must be offered by ML-based AIs to provide individuals with sufficient information to support their tasks. However, they further highlight that this degree fundamentally depends on the supported tasks:

> *"If you want your car's camera to recognize traffic signs, the model's comprehensibility doesn't really matter as you don't have time to understand it anyway while driving. But if you have a model that tells you whether a customer is likely to churn or not, then you want to know why. If the customer is likely to churn, is it because the customer pays too much or because the customer got some bad support? What kind of activities can you take to keep the customer from churning? Then it's all about comprehensibility." (i18)*

Hence, experts state that organizations must understand how much information ML-based AIs must provide about how they produced their output to equip individuals with sufficient information to support their tasks. If ML-based AIs cannot provide required information, their support likely becomes useless as it is the case in the above quote's example. In the worst case, ML-based AIs may even prevent individuals from conducting their tasks if the ML-based AIs hinder them from accessing required information. A bad fit between ML-based AIs' comprehensibility and tasks' required information may thus harm individuals' effectiveness. They further highlight that an ML-based AI's comprehensibility does not only include to render their working principles appropriately transparent, but further comprises the interpretability of its output's content and quality. If individuals fail to comprehend ML-based AIs' output correctly, they likely base their tasks on wrong assumptions, leading to individuals using output in the wrong way as part of their task execution. Moreover, if individuals fail to interpret ML-based AIs' quality measures correctly because they are presented in a format that is not understandable to them, their ability to evaluate an ML-based AIs' trustworthiness may become limited, potentially preventing them from recognizing ML-based AIs' erroneous or unfair behavior:

> *"An AI may tell the user: 'This is now a 90% probability'. But how does the user know what that means? In the end, there may have to be a traffic light or something like that – but that always makes users believe that there is a certain level of reliability, which may not even be there." (i16)*

Thus, the experts warn that wrong interpretations of both AIs' output and quality measures may lead to individuals adopting AIs' wrong or unfair behavior in their task execution.

Production timeliness. Goodhue and Thompson (1995) already proposed production timeliness as a fit characteristic in their original TTF paper. In our interviews, it got apparent that it also constitutes a key characteristic for ML-based AI which can have a significant impact on the other fit factors (as we will discuss in the next paragraph). Depending on the input data's volume that has to be processed and the data volume and algorithm(s) used to create an ML-based AI, its latency when being used can vary significantly (Cheng et al., 2016). The experts therefore highlight that the timeliness of ML-based AIs' support has to be aligned with the required latency of the supported tasks. Slow AIs may slow down individuals' task execution when individuals have to wait for the AIs' responses to base their tasks on their outputs. Moreover, if AIs fail to act within time frames required by their supported tasks, their produced outputs may become useless for the individuals, as demonstrated by the following quote:

> *"What's always a major issue: When is an AI's latency really helpful? This strongly depends on the use case. One may say 'I have offshore wind turbines. This means that I need to know about any damages three months in advance to be there in time.', while another says 'I'm in the production hall and can react within ten seconds. Thus, it would be enough if the AI predicts any damage within twenty seconds.'" (i3)*

As highlighted by the experts, organizations must therefore understand the required production timeliness of individuals' tasks to align AIs accordingly. Otherwise, their individuals may not benefit from the AIs' support which may even harm the individuals' efficiency and may even hinder task execution.

Cross-characteristic dependencies. Lastly, the experts strongly stress that interrelations between the different characteristics have to be considered when assessing task-AI fit. They highlight that, to adjust single task-AI fit characteristics, it is usually necessary to alter used algorithms or data. As a result, organizations often have to face resulting trade-offs between different task-AI characteristics. For example, literature and experts stress that, while both fairness and reliability can be controlled by letting the AI focus more on specific aspects, related adjustments may create a dilemma as changes to reduce certain errors may create unfair behavior in other aspects or even decrease predictive performance (Corbett-Davies and Goel, 2018). Production timeliness may render overly complex and slow AIs insufficient and may require organizations to trade faster reacting AIs against more reliable, fair, and comprehensible ones if faster algorithms support these issues less well (Russell and Norvig, 2016). While complex algorithms may result in higher predictive performance (Kaplan et al., 2020), their complex operating principles may reduce their comprehensibility (Miller, 2019), potentially forcing organizations to sacrifice comprehensibility for less erroneous and unfair behavior of ML-based AIs:

> *"Maybe this approach is five percent less reliable than neural networks, but it at least allows you to comprehend why something happens. If it is less relevant that a human can comprehend what happens, then I can go with neural networks and trade high comprehensibility with higher reliability. But then, I give up that certain things can be understood." (i6)*

Thus, we posit:

Proposition 5: *In the context of ML-based AI, reliability, fairness, comprehensibility, and production timeliness likely are the central task-AI fit characteristics that impact individual performance.*

Proposition 6: *In the context of ML-based AI, cross-characteristic dependencies likely cause tradeoffs between task-AI fit characteristics that impact individual performance.*

4.5 Discussion

In this study, we examined the relationship between ML-based AI use and individual performance. We developed a theoretical model for this linkage, which has not yet been studied on an abstract level in IS research. Due to our study's explorative nature, we followed a qualitative research approach. We used data from 24 expert interviews and AI literature to deduct a theoretical model that can be used for empirical research. Building on the widely used

TTF model, we developed dimensions for the TTF constructs before expanding it with new components to fit the AI context. In detail, we added the *data characteristics* and *task-data fit* constructs as data availability and quality largely determine AI technology's suitability for given tasks according to literature and our experts. *Task-data fit* and *AI characteristics* then determine *task-AI fit*, i.e., the match between AI particularities and given tasks. According to our analysis, *task-AI fit* should be the main predictor for *utilization* and *individual performance*.

Our study makes several theoretical contributions. Besides implementations for specific use cases (e.g., Kumar et al., 2018; Liebman et al., 2019), IS research on ML-based AI has so far mostly focused on user interaction with AI systems (Rzepka and Berger, 2018) and ethical considerations, such as transparency (e.g., Chai and Li, 2019; Fernandez and Provost, 2019) or fairness (e.g., Haas, 2019; van den Broek, 2019). To the best of our knowledge, we are among the first to study the linkage between ML-based AI use and performance impacts, thus answering a call for research regarding human-AI hybrid systems (Rai et al., 2019). We propose a theoretical model based on a rigorously conducted qualitative research approach that explains performance gains through AI use as a function of task, data, and technology characteristics. In addition, we conceptualize the main theoretical constructs using data from our expert interviews by identifying the most relevant subdimensions for each construct. Thus, we enable empirical testing of our model in various contexts where ML-based AI might be applied to support humans. Although we focused on individual performance, the proposed model should be transferable to group- or even organizational-level analyses of TTF-related performance impacts. Our results also confirm the TTF model's flexibility, which has already been applied in a variety of contexts ranging from GSS (e.g., Zigurs and Buckland, 1998) to mobile IS (e.g., Gebauer et al., 2010). Our study's findings also comprise significant contributions for practitioners. The reasoning behind our model can be used to validate possible initiatives to introduce ML-based AI for specific use cases. In detail, decision-makers can examine characteristics of tasks, data, and available AI technology to estimate fit and subsequently performance impacts for given use cases. Going back to the examples from the introduction, physicians could, e.g., identify data availability as the main challenge for applying AI for medical diagnostics successfully (e.g., due to privacy concerns) and bankers could assess comprehensibility to be the central issue in the credit scoring context (e.g., due to regulatory requirements). As the model is built on diverse experience of experts from practice, we can assume its applicability for a variety of industries.

Of course, our study is subject to some limitations. First, we did not perform empirical testing of the proposed model. Here, future studies should focus on the perspective of affected individuals to allow evaluating the impact on individuals directly. This is also needed to verify whether the corresponding user perspective is sufficiently represented, as our model is mostly based on a managerial and IT professional perspective due to the interviewees' background. Second, although we aimed to cover many industries and use cases when selecting interviewees, we cannot eliminate potential data biases towards specific industries completely. Again, quantitative studies in varying contexts should help to uncover such biases to validate the model's applicability.

5 Paper C: The Impact of CV Recommender Systems on Procedural Justice in Recruiting

Title

The Impact of CV Recommender Systems on Procedural Justice in Recruiting: An Experiment in Candidate Selection

Authors

- Verena Eitle, Technical University of Darmstadt, Germany
- Felix Peters, Technical University of Darmstadt, Germany
- Andreas Welsch, Technical University of Darmstadt, Germany
- Peter Buxmann, Technical University of Darmstadt, Germany

Publication Outlet

Proceedings of the Twenty-Ninth European Conference on Information Systems (ECIS 2021), A Virtual AIS Conference

Abstract

Due to the increasing amount of digitally available applicant information recruiters have difficulties to manage applications through manual recruiting practices. Using CV recommender systems in the selection phase supports recruiters in identifying the most suitable candidates by computing the similarity between a candidate’s profile and job requirements. While recent research has mainly focused on technical improvements, we seek to gain more insights about human-algorithm interactions in recruiting. Our study aims to examine what impact the use of a CV recommender system has on procedural justice in the selection process. Through an experimental set-up with 74 recruiters from 22 multinational companies, our study shows that the incorporation of a CV recommender system helps recruiters to ensure the rule of consistency and bias suppression in the selection phase. Thus, our quantitative results indicate that CV recommender systems can have an impact on procedural justice in candidate selection.

Keywords

Candidate Selection, CV Recommender Systems, Procedural Justice

5.1 Introduction

Advancements in information systems and social developments have significantly influenced the way of working in the field of human resource management (HRM). In recent years, organizations have shifted their priorities towards HRM as they perceive their workforce as one of their most important assets. The increasing demand for qualified talents might also result in a war for talents as the shortage of talents is considered one of the most worrying concerns among CIOs and IT executives in 2019 (Kappelman et al., 2020). To attract, select, and retain these talents, recruiting has become a strategic priority in organizations. Black and van Esch (2020) argue that digitization has made a major contribution to further developments in recruiting and emphasize the following eras of e-recruitment. Digital Recruiting 1.0 and 2.0 enable organizations to post job openings on digital job boards on the internet and social network platforms such as LinkedIn. Organizations have the opportunity to narrow down their target group of potential candidates and to contact them directly with concrete job postings (Black & van Esch, 2020). By searching through many digital job postings with a few simple clicks, potential candidates are able to submit multiple applications with less effort. As a result, the increase of incoming applications has made the manual recruiting process more difficult for organizations as recruiters have to manually process digitally available applicant information. While coping with this large amount of applications, recruiters also need to ensure fairness in the selection process as their decision has a major impact on the applicants future (Arvey & Renz, 1992; Gilliland, 1993). However, procedural justice along the decision-making process in the selection phase is often impeded by recruiters' previous work experiences, own beliefs or personal biases (Åslund & Skans, 2012; Eckhardt et al., 2014).

To cope with the increasing amount of data and to ensure fairness, different types of artificial intelligence (AI) technologies have been integrated into the recruiting process (Strohmeier & Piazza, 2015; van Esch et al., 2019) which Black and van Esch (2020) describe as Digital Recruiting 3.0. In particular, the development of Curriculum Vitae (CV) recommender systems is an essential research area in the selection phase of recruiting. These systems are typically applied in the selection phase of the recruiting process (Schneider, 1987) to estimate the person-job (P-J) fit (Caldwell & O'Reilly, 1990; Edwards, 1991; Kristof-Brown, 2000; Wilk & Sackett, 1996). By computing the similarity between the details of a candidate's profile and the given job requirements, CV recommender systems can support recruiters in identifying the most suitable candidates. While the performance level is constantly increasing due to technical improvements (e.g., Bansal et al., 2017; Lu et al., 2013; Malinowski et al., 2006), little is known about the socio-technical context of the interaction between human recruiters and CV recommender systems (Green & Chen, 2019). Since the final decision in candidate selection still remains in the power of recruiters, further insights about the human-algorithm interactions are essential in order to investigate the effect on procedural justice. Therefore, our study aims to examine what impact the use of a CV recommender system has on procedural justice in the selection process. As a research design, we have chosen an experimental set-up in which 74 recruiters from 22 large multinational companies were given the instruction to create top-10 rankings of candidates for two fictional job postings. By randomly assigning the participants to either the control group which represents the non-CV recommender system supported settings or to the treatment group in which recruiters received a matching score generated by a CV recommender system, we were able to investigate our research question. Our study contributes to research and practice in the field of recruiting by providing quantitative findings that CV recommender systems tend to ensure procedural justice as recruiters are able to rank candidates in a more consistent manner and are more likely to assess a candidate's knowledge, skills, and abilities when relying on the CV recommender system.

The rest of this paper is structured as follows: Section 2 outlines the theoretical background of recommender system in recruiting with a focus on CV recommender systems and elaborates on procedural justice in candidate selection. After describing the research design in the form of an experimental set-up in section 3, we present the results of the quantitative study in section 4. The discussion, the contributions to research and practice as well as the limitations and opportunities for future research are outlined in section 5, followed by the conclusion in section 6.

5.2 Theoretical Background

5.2.1 Overview of Recommender Systems

Over the last couple of years, the overload of information with which people need to cope on a daily basis has resulted in complex decision-making environments. The fact that humans have difficulties making decisions due to their limited cognitive resources and time constraints in evaluating and processing available information was coined by Simon (1955) as the phenomenon of bounded rationality. In order to help people deal with the overwhelming amount of data and to support them in the intelligence, design, choice, and implementation phase of complex decision-making processes (Simon, 1960), recommender systems have been developed. By generating personalized suggestions, recommender systems offer only a small number of selection options and eliminate irrelevant and excessive information (Adomavicius

& Tuzhilin, 2005; Burke, 2002). To be more precise, the primary use of recommender systems is to predict elements that a user is likely to evaluate as positively according to his or her underlying preferences (Ricci et al., 2011). In general, recommender systems can be classified into the following categories: Content-based, collaborative filtering, and knowledge-based recommender systems (Aggarwal, 2016; Burke, 2002; Ricci et al., 2011). Content-based recommender systems recommend items to users that are similar to those that they have historically favored or expressed interest in. In order to retrieve a user's preferences, tastes, and desires, the recommender system uses long-term user profiles with user attributes that have been accumulated over time. By matching these user attributes to item attributes, new items will be recommended to the user (Adomavicius & Tuzhilin, 2005; Aggarwal, 2016; Pazzani & Billsus, n.d.). Since contentknowledge is mainly derived from unstructured or semi-structured data, item descriptions are composed of a set of textual features that can be acquired by various information retrieval or information extraction methods with the help of statistical, machine learning, or natural language processing techniques (Lops et al., 2011). In contrast, collaborative filtering recommender systems generate item recommendations based on the similarity towards other users' preferences (Adomavicius & Tuzhilin, 2005; Aggarwal, 2016; Schafer et al., n.d.). This type of recommender system has to cope with a so-called cold-start issue as a new user has to first rate several items or a new item has to receive a couple of ratings before a user similarity can be determined (Bobadilla et al., 2013; Ramezani et al., 2008). Recommendations generated through knowledge-based recommender systems are derived from specific domain knowledge which have to be acquired through interviews or other knowledge discovery techniques (Aggarwal, 2016). A common form of knowledge representation are ontologies which display relations among attributes, objects, and item features. The main downside of this recommender system lies in the high efforts of knowledge acquisition (Ramezani et al., 2008).

5.2.2 Recommender Systems in Recruiting

According to the attraction-selection-attrition (ASA) framework by Schneider (1983), organizations tend to achieve a certain degree of homogeneity among their employees by identifying candidates during the recruiting phases of attraction, selection, and attrition who have similar characteristics and behaviors as the organization. The empirical study by Judge and Cable (1997) revealed that in the attraction phase, potential candidates search for suitable job postings and organizational cultures based on their own personality, preferences, and field of interest. Particularly in the attraction phase, there is a tendency of organizations to achieve a certain degree of homogeneity by seeking to recruit candidates with similar attributes and behaviors which is also described by the term "right types" (Schneider, 1983). In the selection phase, organizations seek to select candidates who possess specific competencies and skills required for the job position. By narrowing the applicant pool using pre-selection techniques and face-to-face interviews, companies are able to select a homogeneous group of candidates with specific skills (Bretz et al., 1989; Schneider, 1983). During the attrition phase, there is a tendency for employees who do not fit into the organization to eventually leave, while employees who embrace the organizational culture strive to retain their jobs and pursue their careers over time (Chatman, 1989; Schneider, 1983). By retaining the "right types" in the organization who share similar characteristics and behaviors, companies can increase the homogeneity among their workforce (Schneider, 1983). Since the increase in digital job and applicant data particularly impedes the screening and assessment activities of recruiters (Black & van Esch, 2020), the following sections mainly refer to the selection phase in recruiting.

The main task in the selection phase is the matching of potential candidates and job postings, which is an essential subject of the person-job (P-J) fit and person-organization (P-O) fit literature (Adkins et al., 1994; Judge & Cable, 1997; Rynes & Gerhart, 1990; Wilk & Sackett, 1996). The overarching research relates to the fit between a person and the environment, which has been a pervasive component in major research areas including personality theory, occupational psychology, personnel selection, and social psychology (Schneider, 2001). According to the person-environment fit concept (P-E), behavior is influenced by the congruence between personal and situational variables and not just by one of the elements alone. To be more precise, the compatibility between personal variables including abilities, needs, and values as well as environmental variables such as organizational culture, task demands, and job attributes leads to either positive or negative outcomes (Kristof, 1996; Muchinsky & Monahan, 1987; Ostroff, 1993; Schneider, 2001). Besides the P-J and P-O fit, the comprehensive P-E fit concept comprises further sub-categories including the person-vocation (P-V) fit, the person-group (P-G) fit, and the person-supervisor (P-S) fit (Kristof-Brown et al., 2005; Sekiguchi, 2004).

In the recruiting literature, the concepts of P-J and P-O fit predominate the selection phase of Schneider's (1983) ASA framework since the primary objective is to match individuals and jobs. The operationalization of the P-J fit by Edwards (1991) refers to the demands-ability fit and the needs-supplies fit. To be more precise, the demands-ability fit determines the extent to which an employee's knowledge, skills, and abilities, the so-called KSA’s, meet the requirements of a job. These KSA’s comprise, for example, work experience, technical skills, problem-solving skills, academic experience, and leadership skills (Kristof-Brown, 2000). The needs-supplies fit, on the other hand, addresses whether needs, wishes, or preferences of an employee are satisfied by the jobs’ characteristics and attributes (Edwards, 1991; Kristof, 1996; Sekiguchi, 2004). However, since candidates tend to select the vacant job positions according to their own needs and preferences (Judge & Cable, 1997), the primary task of recruiters is to identify candidates with the required KSA’s. The study by Caldwell and O’Reilly (1990) showed that the match between the KSA’s of a candidate and the job requirements positively influences an employee's job performance and ultimately job satisfaction. Furthermore, Wilk and Sackett (1996) reported that the match between an employee's skills and the complexity of the job even allows the employee to move up in the job hierarchy in the future. These empirical results indicate that the demands-ability fit is crucial for assessing the P-J fit (Kristof-Brown, 2000). With regard to the operationalization of the P-O fit, Chatman (1989) argues that the congruence between candidates' values as well as organizational norms and values can have a positive impact on the selection phase since this match increases the likelihood that a candidate identifies himself with the organizational culture. An experiment conducted by Kristof-Brown (2000) revealed that recruiters explicitly distinguish between the P-J fit and the P-O fit when selecting applicants. When assessing the first group of applicants, recruiters tend to follow the P-J fit as they primarily consider the KSA’s as their main selection criteria. In the subsequent evaluation rounds of the recruiting process, the emphasis is on the P-O fit since the match between personal values and organizational values is given higher priority (Kristof-Brown, 2000; Rynes & Gerhart, 1990).

With the advancements of Digital Recruiting 1.0 and 2.0 (Black & van Esch, 2020), which allow organizations to post their job openings on digital job boards and professional and social networking platforms like LinkedIn, the amount of digital candidate data has increased significantly. Since the selection phase involves a high proportion of manual tasks, managing the large volume of digital applications can be time-consuming and costly for organizations (Eckhardt et al., 2014; Strohmeier & Piazza, 2015). While different types of artificial intelligence (AI) technologies can be integrated throughout the recruiting process (Strohmeier & Piazza, 2015; van Esch et al., 2019), the emergence of recommender systems have particularly simplified the manual tasks of recruiters in the selection phase. The study by Faerber et al. (2003) has compared the prediction performance of a content-based recommender system, a collaborative filtering recommender system, and a hybrid approach in the field of CV recommendations. According to their findings the content-based approach yields the best results in matching a candidate's profile and the job requirements. Based on the P-J fit, Malinowksi et al. (2006) have developed a CV recommender system that follows the demands-ability fit approach (Edwards, 1991) by recommending candidates whose CVs most closely match the specific job requirements. In order to address the needs-supplies fit approach (Edwards, 1991) by matching a candidate's preference with the job attributes, the authors additionally developed a job recommender system. Based on a latent aspect model both recommender systems are able to compute the similarity between the candidate’s profile and the job requirements. Since the predictive quality of the two recommender systems was the main subject of the study, the computer-generated recommendations were compared with the original list of jobs selected by the study participants and the original list of top candidates. The results showed that the predictions of the CV and the job recommender systems largely corresponded to human choices, which indicate a high prediction quality and promising system performance. Moreover, the contentbased recommender system proposed by Lu et al. (2013) is designed as a hybrid model that integrates a CV and a job recommender system in one system. The profile-based similarity of the candidate’s details and the job posting was computed by using latent semantic analysis (LSA) tools. In addition, the recommender system is capable of not only including a candidate's profile and the job requirements, but also processing user interactions. In an experiment, the participants were able to indicate their preferences through the interaction features "posted", "applied", "favorited", "liked", and "visited". The study by Almalis et al. (2015) extents the research of content-based CV recommender systems in a way that the match between human KSA’s and job attributes is based on different value ranges, such as specific values, a range with lower limit, a range with upper limit, and a range with both lower and upper limit. In other words, the proposed CV recommender system is able to differentiate between job requirements that refer to ranges of values such as “candidates must be at least 40 years old” or “between 18-40 years old”. The proposal of a further hybrid content-based recommender system by Bansal (2017) facilitates the matching of candidates profiles and job postings from the perspective of recruiters and job seekers in an integrated system. Instead of using words as textual features, the researcher focused on topic features by applying the topic modelling algorithm Latent Dirichlet Allocation (LDA). Since this unsupervised machine learning technique allows to detect latent topics that are hidden in the text corpus, low-frequency terms can become quite significant as they are linked to other high-frequency terms.

As shown, current research in the selection phase focuses mainly on enhancing the prediction performance of CV recommender systems by evolving algorithms and improving technical features (Almalis et al., 2015; Bansal et al., 2017; Färber et al., 2003; Lu et al., 2013;

Malinowski et al., 2006). However, instead of optimizing computational performance, Green and Chen (2019) emphasize that attention in research should rather shift towards a socio-technical context to explore how humanalgorithm interactions can be improved. According to their algorithm-in-the-loop framework, algorithmic aid can help to improve the decision-making process by incorporating algorithms which inform and advise humans in their decision-making while the final decision still remains with humans. Although the study of human-algorithm interaction is developing slowly in areas such as web journalism (Christin, 2017), forecasting (Dietvorst et al., 2018), and criminal justice (Green & Chen, 2019; Grgić-Hlača et al., 2019), research has not yet sufficiently taken into account the socio-technical context in the field of recruiting (Green & Chen, 2019; Grgić-Hlača et al., 2019).

5.2.3 Procedural Justice in Candidate Selection

Despite the fact that key performance indicators in the recruiting process are largely standardized, the selection process for candidates often differs among recruiters. Previous work experience, individual attitudes, and personal preferences lead to a variety of different behaviour patterns among recruiters which can significantly influence the selection of suitable candidates (Eckhardt et al., 2014). Furthermore, the existence of conscious or unconscious cognitive bias among recruiters might also contribute to the likelihood of inconsistent decision-making processes in the selection phase (Åslund & Skans, 2012; Black & van Esch, 2020). These diverse set of behaviour patterns among recruiters increase the risk of unfairness in the selection phase and can ultimately compromise a candidate's chance of being selected.

In order to examine fairness in the decision-making process during the selection phase, the literature on organizational justice (Greenberg & Colquitt, 2005) must be taken into account which primarily addresses employees' reactions regarding unfairness and inequity in an organizational context and distinguishes between distributive and procedural justice. Distributive justice describes the degree to which an employee perceives the distribution of outcomes such as payments and rewards as fair in the sense of equity (Adams, 1965; Cohen, 1987) and equality (Deutsch, 1975). When considering the equity principles which determine the distribution of resources according to the contributions of employees, the foundation of distributive justice refers to Adam's (1965) equity theory. In contrast, procedural justice refers to the perceived fairness in the actual decision-making process that ultimately determines the outcome (Greenberg & Colquitt, 2005). In order to ensure that the procedure can be assessed as fair, Leventhal (1980) defined the following six rules for procedural justice: consistency, unbiased suppression, accuracy, correctability, representativeness, and ethicality. Since fairness in the candidate selection process depends on procedural justice, Gilliland (1993) and Arvey and Renz (1992) have defined specific procedural rules for the selection phase. In this context, the rule of consistency should be emphasized as Leventhal (1980) and Gilliland (1993) recommend a certain degree of uniformity in the selection procedure since all candidates should have the chance to receive the same decision-making process regardless of demographics, personality, or background. Arvey and Renz (1992) point out that consistency in candidate selection is only given when the content of the selection system, the scoring, and the interpretation of scores are standardized across all applicants. In addition, the rule of bias suppression by Leventhal (1980) is also crucial to ensure procedural justice in the selection phase as it determines that recruiters should not make decisions based on their own self-interest or be influenced by their own beliefs and opinions (Leventhal, 1980). To ensure objectivity rather than risking subjectivity, Arvey and Renz (1992) suggests to apply quantifiable methods which take certain criteria into account rather than relying on the recruiters' instincts and experiences. The suppression of personal bias is also addressed in the propriety of questions as improper questioning and prejudicial statements impede the level of fairness in the selection phase (Gilliland, 1993).

By examining procedural justice in e-recruiting tools, the findings of Thielsch et al. (2012) show that applicants expect a higher level of objectivity when using an e-recruiting tool compared to traditional manual recruiting practices. In addition, the qualitative study by Ochmann and Laumer (2019) proposes that the implementation of AI-based instruments could contribute even more to increase the level of fairness by increasing objectivity during the selection phase. While traditional selection methods have been perceived as unfair due to the risk of personal bias on part of the recruiters, the qualitative findings suggest that AI technologies could assist in the decision-making process by focusing solely on the candidates' skills and thus increasing objectivity. It should be noted, however, that the level of user reliance in a technology is also considered a critical factor in achieving procedural justice, as the final selection decision still remains in the power of recruiters. Reliance towards a technology depends primarily on user acceptance and the degree of influence that the user allows in their judgment (Arnold & Sutton, 1998; Madhavan & Wiegmann, 2007). Following the study by Ötting and Maier (2018) which empirically examined the impact of human and AI-based intelligent systems on procedural justice in a generic work-life situation, we aim to gain empirical insights into procedural justice in the selection process.

Based on the outlined literature on candidate selection (e.g., Caldwell & O'Reilly, 1990; Edwards, 1991; Schneider, 1987; Wilk & Sackett, 1996), CV recommender systems (e.g., Almalis et al., 2015; Bansal et al., 2017; Malinowski et al., 2006), and procedural justice (Arvey & Renz, 1992; Gilliland, 1993; Greenberg & Colquitt, 2005), we believe that incorporating a CV recommender system in the selection phase could increase procedural justice by helping recruiters to ensure consistency and objectivity in their decision-making process. Under the premise that recruiters rely on a CV recommender system and take the generated suggestions into account, we anticipate that the top-10 rankings of recruiters who incorporate a CV recommender system into their decision-making process will be more consistent and similar than the top-10 rankings of those who rely solely on their own judgement without using a CV recommender system. Furthermore, we would like to gain further insights into the demands-ability approach in the context of the P-J fit (Edwards, 1991; Kristof-Brown, 2000) when incorporating a CV recommender system in the decision-making process of recruiters. As outlined above, CV recommender systems are based on the demands-ability fit as they compute the similarity between the applicant's KSA's and the respective job requirements (Almalis et al., 2015; Bansal et al., 2017; Färber et al., 2003; Lu et al., 2013; Malinowski et al., 2006). Since the suggestions generated by CV recommender systems are based on the KSA's of candidates and are not exposed to subjective discrimination or personal bias by recruiters, we anticipate that the CVs of the top-10 ranked candidates which were selected with the help of a CV recommender system possess stronger KSA's than those ranked on the basis of the recruiters' sole judgment.

5.3 Methodology

Since the aim of our study is to examine what impact the use of a CV recommender system has on procedural justice in the selection process, we conduct an true experimental research with a posttest-only control group that enables us to determine cause-effect relationships (Campbell & Stanley, 1963; Gay et al., 2012). The research design of the experiment is illustrated in Figure 7 and is described in detail in the following section.

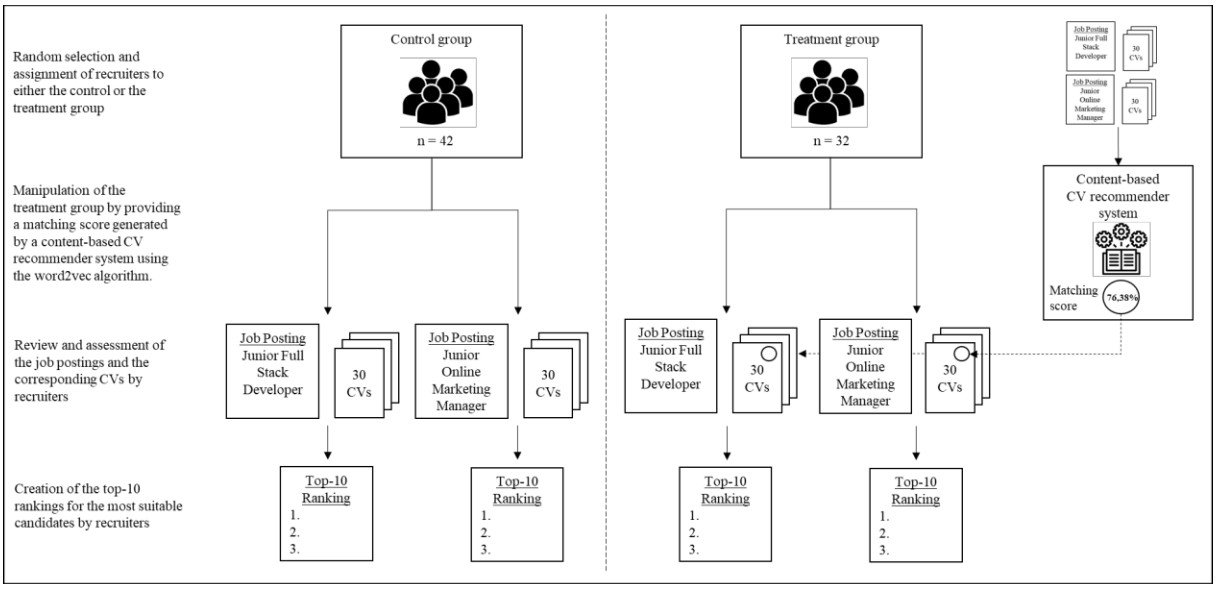
Random selection and assignment of recruiters to either the control or the treatment group
Manipulation of the treatment group by providing a matching score generated by a content-based CV recommender system using the word2vec algorithm.
Review and assessment of the job postings and the corresponding CVs by recruiters
Creation of the top-10 rankings for the most suitable candidates by recruiters
Control group
n = 42
Treatment group
n = 32
Job Posting Junior Full Stack Developer
Job Posting Junior Online Marketing Manager
30 CVs
Content-based CV recommender system
Matching score
76,38%
Top-10 Ranking
1.
2.
3.

Figure 7: Experimental research design

In order to establish a realistic experimental set-up for a decision-making process in the selection phase, we have involved professional recruiters rather than non-professional study participants. Through a cooperation with a national association for employer branding, talent marketing, and recruiting, we were able to randomly select recruiters who were willing to participate in our experiment. The random selection method is recommended primarily because it ensures external validity by increasing the degree to which the study results can be generalized to other groups (Campbell & Stanley, 1963; Dean et al., 2017; Kirk, 2012). By following a between-subject design, we have also applied the randomization method when assigning participants to either the control or the treatment group. Random assignment is particularly needed to ensure internal validity as it reduces systematic bias between the treatment and the control group by distributing participants equally among these groups (Campbell & Stanley, 1963; Kirk, 2012; Mikolov et al., 2013). The manipulation of the independent variable refers to a matching score generated by a CV recommender system and distinguishes the groups as follows: The control group represents the non-CV recommender system supported setting in which the participants have received CVs without any suggestions generated by a CV recommender system. The treatment group represents the CV recommender system supported setting in which the participants have received the same CVs but to which a matching score generated by the applied CV recommender system has been added in the upper right corner.

Following the current research on recommender systems in the field of candidate selection (Almalis et al., 2015; Bansal et al., 2017; Färber et al., 2003; Lu et al., 2013; Malinowski et al., 2006), we have decided to also use a content-based CV recommender system that supports recruiters in the selection phase to identify suitable candidates. As we seek to examine the effect of CV recommender systems on procedural justice rather than improving the performance level through technical advancements, we decided to use an existing CV recommender system developed by a global enterprise software provider with sufficient training data. The underlying machine learning technique refers to the word2vec algorithm by Mikolov et al. (2013) which represents a neural network model with a single hidden layer. In the case of the applied CV recommender system, data cleansing activities such as functional removal, lower case, and plural removal are performed on the input document in an initial step. After this prerequisite is fulfilled, the input document is tokenized into corresponding bigrams and trigrams. By using the word2vec algorithm, each token is assigned to a word embedding which ultimately represents a vector space. In order to remove irrelevant tokens and to generate interpretable token clusters, the tokens in the form of word embeddings are assigned to certain branches of a formerly created skill tree. As the final goal is to compare a CV and a job posting, the word embeddings are combined into document-level embeddings to compute the cosine similarity between the document vectors. The output of the selected content-based CV recommender system is a matching score which is expressed as a floating-point number between 0.0 and 1.0. The higher the value of the matching score, the closer the similarity between the CV and the job profile, and the higher the rank of the CV in a list of suitable candidates.

In regard to the experimental set-up, we created two fictional job posting based on examples from the participating companies: one for a Junior Full Stack Developer and one for a Junior Online Marketing Manager. We focused on junior positions as these positions oftentimes receive large numbers of applications and are thus more attractive for the implementation of a CV recommender system. As a second step, we collected a diverse set of 30 CVs for each job posting from computer science, information systems, business, and marketing students of higher education institutions. During the experiment, the participants received the task description through a survey tool in which the two job postings as well as the corresponding CVs were available in the form of PDF documents. Based on the random assignment to either the control group or the treatment group, the CVs either included a matching score generated by the CV recommender system or not. According to the task description, the participants of both experiment groups were instructed to first read the job postings and the corresponding CVs thoroughly. Based on a careful assessment of the candidates and the requirements of the first job posting, the recruiters were asked to create a ranking in the survey tool based on the suitability of the candidates under the assumption that the top-10 ranked candidates would be invited for a further interview. Subsequently, the participants were encouraged to proceed with the creation of the ranking list for the second job posting. Since the ranking represents the final outcome of the experiment and is considered in the further data analysis, our study is designed as posttest-only control group. This approach allows us to avoid testing effects that could have had an impact on the participants' behaviour if they were exposed to any kind of information in advance (Gay et al., 2012).

5.4 Results

Regarding the participation rate, 89 professional recruiters voluntarily signed up for our experiment, out of which 74 completed the tasks (83% response rate). At the time of the experiment (January 2019), these 74 participants were employed in 22 large multinational companies. Among all participants, 74% were female, 83% were between 25 and 44 years old, and 78% had at least three years of experience in recruiting. To ensure objectivity within the quantitative data analysis, we manually extracted variables from all CVs in a two-stage procedure. First, variables were extracted independently by two of the authors. Subsequently, results were synchronized to reach consent and to apply consistent standards. In accordance with the P-J fit which determines the suitability between the KSA's of candidates and the job requirements (Edwards, 1991; Kristof-Brown, 2000), we extracted the following variables from all CVs: study duration (in years) and relevant working experience (in years). By conducting statistical tests, we were able to relate these variables to the observed behaviour of the participating recruiters. The data was pre-processed using the Python programming language and subsequently analysed using SPSS. Given the nature of our study, we tested for significance at a 10% level to avoid discarding interesting relationships (Rosnow & Rosenthal, 1989; Schumm et al., 2013). In regard to the following section, we present the results as mean ± standard deviation, unless we state otherwise. While screening the experiment data, we detected three cases in which participants only completed the marketing job posting, and one case in which only the development task was finished. We decided to keep these partial completions in our dataset to account for the rather small sample size.

To examine whether the top-10 rankings of recruiters who were supported with the matching score generated by the CV recommender system are more consistent and similar than those who rely on their own judgement, we first calculated pairwise correlations between the rankings of participants separately for the CV recommender system supported group and the non-CV recommender system supported group (in the following referred to as inner group ranking correlation). Here, the ranked candidates received their respective position, while candidates outside the top-10 were being ranked as 11th , thus creating a lot of ties in our rankings.

Consequently, we chose Kendall's tau as correlation metric for this analysis, as this metric is more robust in the presence of ties in rankings (Kendall, 1946). We then conducted two separate independent-samples t-tests (i.e., one for each job posting) to examine the effects of CV recommender system support on the inner group ranking correlation. The results of our quantitative analysis are summarized in Table 10.

Table 10: Effect of CV recommender system support on inner group ranking correlation

Factor	Task	Levels	Inner group rank. corr.		df	t	Sig.	Cohen's d
			Mean	Std. Dev.				
CV recommender system support	Development	Supported	.440	.233	1235	-21.989	.000	1.281
		Unsupported	.144	.231				
	Marketing	Supported	.489	.274	1300	-18.958	.000	1.057
		Unsupported	.192	.288				
Note: Inner group ranking correlations are calculated as pairwise correlations between rankings from participants of the respective group, as measured by Kendall's tau. Results are based on independent-samples t-tests.								

Based on our quantitative analysis we found statistically significant differences between the CV recommender system supported and non-CV recommender system supported groups with regard to the inner group ranking correlation score. The results showed that the inner group ranking correlation was higher in the CV recommender system supported (development task: .440 ± .233; marketing task: .489 ± .274) than in the non-CV recommender system supported groups (development task: .144 ± .231; marketing task: .192 ± .288). In other words that means that the rankings from recruiters who received the matching score generated by the CV recommender system were more strongly correlated with each other than rankings from recruiters without the CV recommender system support. For both groups, the effects were statistically significant (development task: t = -21.989, p < .001; marketing task: t = -18.958, p < .001) and effect sizes were larger than one standard deviation, as measured by *Cohen's d* (development task: 1.281; marketing task: 1.057).

According to our results, we can strongly support our anticipation that the top-10 rankings of recruiters within the CV recommender system supported group are more consistent and similar than those who did not received any matching score from the CV recommender system. We can further suspect that recruiters relied on the matching score generated by the CV recommender system. To further examine this finding, we also calculated the average correlation between the recruiters' rankings and the ranking proposed by the CV recommender system. We found a strong correlation for both tasks (development task: .583 ± .253; marketing task: .599 ± .268), which further supports our assumption. For comparison, in the unsupported groups the observed correlations were much lower (development task: .062 ± .231; marketing task: .080 ± .310).

To examine our second anticipation that the CVs of the top-10 ranked candidates which were selected using a CV recommender system possess stronger KSA's than those which were ranked without any CV recommender system support, we compared the ranked candidates of the control and the treatment group based on the extracted variables of study duration and relevant working experience. We calculated averages for ranked candidates on a per-recruiter basis and then compared between values from both groups using independent-samples t-tests. Once again, we considered rankings from development and marketing job postings separately. The quantitative results are summarized in Table 11.

Table 11: Effects of CV recommender system support on KSA levels of ranked candidates

Task	Variable	Group	Mean	Std. Dev.	df	t	Sig.	Cohen's d
Development	Study duration	Supported	4.909	.324	69	-1.532	.130	.365
		Unsupported	4.792	.311				
	Working experience	Supported	3.500	.505	69	-2.277	.026	.540
		Unsupported	3.194	.622				
Marketing	Study duration	Supported	5.063	.264	71	-2.537	.013	.604
		Unsupported	4.924	.190				
	Working experience	Supported	2.054	.308	71	-.670	.505	.154
		Unsupported	1.993	.470				

Note: Study duration and working experience are measured in years and were extracted from the submitted resumes. Results are based on independent-samples t-tests.

By conducting our quantitative analysis we found that top-10 ranked candidates in CV recommender system supported settings displayed statistically significant stronger levels of KSA's than top-10 ranked candidates in non-CV recommender system supported settings in two out of four observed cases (development – working experience: t = -2.277, p = .026; marketing – study duration: t = -2.537, p = .013). For both cases we observed medium effect sizes (larger than .5), as measured by *Cohen's d* (development – working experience: .540, marketing – study duration: .604). In addition, we found a small effect size (larger than .2) for study duration in the development task, that was not statistically significant (t = -1.532, p = .130, d = .365). Considering these findings, we can partially support our anticipation that candidates of the top-10 rankings possess stronger KSA's in the cases when recruiters have been supported by the CV recommender system compared to the cases were recruiters have not received a matching score generated by the CV recommender system.

5.5 Discussion

To cope with the increasing amount of digital applicant data (Strohmeier & Piazza, 2015; van Esch et al., 2019) and to ensure fairness in the candidate selection phase (Gilliland, 1993; Ochmann & Laumer, 2019; Thielsch et al., 2012), research has increasingly focused on the development of CV recommender systems. These systems serve to identify the most suitable candidates for a given job by calculating the similarities between candidate profiles and job requirements. Thus, CV recommender systems are typically applied in the selection phase of the recruiting process (Schneider, 1983) with the purpose of estimating the P-J fit (Adkins et al., 1994; Judge & Cable, 1997; Rynes & Gerhart, 1990; Wilk & Sackett, 1996). While prior research on CV recommender systems has mainly focused on improving the performance of CV recommender systems on a technical level (e.g., Bansal et al., 2017; Lu et al., 2013; Malinowski et al., 2006), our study addresses the socio-technical context by concentrating on the interaction between the human recruiter and the algorithm (Green & Chen, 2019; Grgić-Hlača et al., 2019). In detail, we examine what impact the use of a CV recommender system has on procedural justice in the selection process. Therefore, we conduct an experiment with 74 professional recruiters from 22 multinational companies, where the task is to create top-10 rankings of candidates for two fictional job postings. According to our quantitative data analysis, we found statistically significant differences between the control and the treatment group with regard to the inner group ranking correlation score. We derive two main findings from our quantitative analysis. First, the analysis of our experiment indicates that the rankings correlated more strongly with each other when recruiters received the matching score generated by the CV recommender system than in the non-CV recommender system supported group. Since this stronger correlation is an indicator that the top-10 ranking list is more consistent and similar among the recruiters of the CV recommender system supported group, we can assume that the level of procedural justice increases through the assistance of the CV recommender system. Second, our quantitative results indicate that the CVs of the top-10 ranked candidates of the CV recommender system supported group contain stronger KSA's in regard to working experience for the development job posting as well as in regard to study duration for the marketing job posting compared to the non-CV recommender system supported group. Due to the presence of these stronger KSA's in the top-10 rankings, our results indicate that KSA's are given more attention when creating the top-10 rankings with the support of a CV recommender system than if recruiters would make the decision on their own. In other words, CV recommender systems can help recruiters to base their decision-making on the pure set of KSA's, rather than being influenced by their own judgment or personal biases. Thus, if candidates possess KSA's required for a particular job posting and a CV recommender system is incorporated in the selection phase, the likelihood of these candidates being selected in the top-10 rankings tends to increase. To summarize, we show that incorporating CV recommender systems increases procedural justice in the selection phase as recruiters are more likely to adhere to the rule of consistency (Arvey & Renz, 1992; Gilliland, 1993; Leventhal, 1980) by ranking candidates in a more consistent and uniform manner. Moreover, we find that the candidates selected by CV recommender system supported recruiters typically possess stronger KSA's than the candidates selected by non-CV recommender system supported recruiters which can be considered as an indicator of ensuring the procedural rule of bias suppression (Arvey & Renz, 1992; Leventhal, 1980).

Our study offers significant theoretical contributions regarding research in the area of human-algorithm interaction. To the best of our knowledge, we are among the first to study the effects of CV recommender system application on procedural justice in the selection phase of the recruiting process. Our study showed that the decision-making process of professional recruiters can be influenced by a CV recommender system by creating more consistent and uniform rankings in which the selected candidates possess stronger KSA's. Thus, we provide quantitative evidence for findings of Thielsch et al. (2012) and Ochmann and Laumer (2019), i.e., that higher levels of objectivity and consistency can be achieved in the candidate selection phase when using an algorithmic aid instead of solely relying on human judgment. Consequently, we show that procedural justice in the selection phase of recruiting can be strengthened by deploying a CV recommender system. We propose that content-based CV recommender systems help recruiters to ensure the procedural rule of consistency (Arvey & Renz, 1992; Gilliland, 1993; Leventhal, 1980) by providing more consistent and uniform rankings. Moreover, relying on these types of systems might mitigate human biases in the recruiting process, such as subjective selection criteria by enabling accurate measurement of candidate's KSA's as proposed by the P-J fit (Caldwell & O'Reilly, 1990; Edwards, 1991; Kristof-Brown, 2000; Wilk & Sackett, 1996).

Our findings also have significant implications for practitioners. We show that organizations should consider deploying CV recommender systems in the selection phase of the recruiting process. Here, the application of such systems might serve several purposes. First, using content-based CV recommender systems increases the likelihood that applicants with higher levels of KSA's will be included in candidate rankings. This way, organizations can ensure that they more strongly consider candidates with a high P-J fit. As a result, more suitable candidates might be identified more efficiently, preventing costly hiring mistakes in the process. Moreover, content-based recommender systems could be used to partly automate the selection process, which would allow to direct further resources towards cognitively more challenging tasks, e.g., estimating the P-O fit via in-person interviews. Second, the deployment of CV recommender systems might reduce existing biases in the recruiting process by making sure that candidate rankings are more consistent across different recruiters.

While our study adds value for both research and practice, it is affected by some limitations that offer opportunities for further research. Despite the fact that we designed our experiment according to realistic recruiting standards and practices by involving professional recruiters, providing real CVs, and using a content-based CV recommender system, we are aware that the experimental set-up has some shortcomings regarding the recruiting process in practice. With regard to the selection phase, our study differs from procedures used in practice to evaluate applicants where CVs are usually reviewed as they are received rather than consecutively in batches. In addition, recruiters would usually receive more information on required skills and context from the hiring manager instead of just referring to the available requirements of the job posting. Furthermore, the choice of a junior job posting could have an impact on the matching score generated by the CV recommender system as the submitted CVs might contain fewer keywords and details than for a professional job posting. Lastly, as we used only one commercially available content-based CV recommender system, we are well aware that the results might differ for alternative solutions.

To increase realism, future research could improve the experimental set-up by providing a centralized CV upload that allows recruiters to review CVs at the time of upload. Therefore, the timeframe of the experiment should also be extended from one to three months in order to make the decision-making process of candidate selection more realistic. In addition, the between-subject design could also be varied by adding another treatment group of recruiters who receive additional information on the key features that influence the generated matching score. Thereby, the effect of increased transparency for recruiters compared to recruiting settings with less information could be studied further. By expanding the research scope with a focus on transparency, additional insights could be gained as to whether recruiters would integrate the suggestions of CV recommender systems even more strongly into their decision-making process as the key features become more transparent. This future research would contribute significantly to the study of human-algorithm interactions in the field of recruiting.

5.6 Conclusion

In recent years, recruiting qualified and skilled talents has gained considerable importance as organizations consider their workforce as strategic assets. While digitization has contributed to the emergence of job portals, recruiters face the challenge of dealing with large amounts of digital applications (Black & van Esch, 2020). In order to cope with this amount of data, CV recommender systems have been developed to support recruiters in the selection phase. By computing the similarity of the candidates KSA's and the job requirements, CV recommender systems are able to identify the most suitable candidates as requested by the demands-ability approach of the P-J fit (Edwards, 1991; Kristof-Brown, 2000). However, relatively little is known about the socio-technical context in which such systems are deployed (Green & Chen, 2019; Grgić-Hlača et al., 2019). Our study aimed to examine the impact of a CV recommender system on procedural justice in the selection phase of the recruiting process. Therefore, we conducted an experiment with 74 recruiters from 22 large multinational companies. Using a between-subject design, we compare top-10 rankings of potential candidates for two fictional job postings between recruiters who are supported by a content-based CV recommender system and unsupported recruiters. Two main observations can be drawn from our quantitative analysis. First, candidate rankings from the CV recommender system supported group exhibit higher levels of similarity than rankings from the non-CV recommender system supported group. Second, candidates selected by recruiters who received the matching score generated by the CV recommender system contain stronger KSA's than candidates selected by recruiters who relied solely on their own judgment. Thus, we find quantitative evidence that the deployment of CV recommender systems can increase consistency and reduce personal bias in the selection phase of the recruiting process, which might improve procedural justice in this process.

6 Paper D: Coordinating Human and Machine Learning for Effective Organizational Learning

Title

Coordinating Human and Machine Learning for Effective Organizational Learning

Authors

- Timo Sturm, Technical University of Darmstadt, Germany
- Jin Gerlach, University of Passau, Germany
- Luisa Pumplun, Technical University of Darmstadt, Germany
- Neda Mesbah, Technical University of Darmstadt, Germany
- Felix Peters, Technical University of Darmstadt, Germany
- Christoph Tauchert, Technical University of Darmstadt, Germany
- Ning Nan, The University of British Columbia, Canada
- Peter Buxmann, Technical University of Darmstadt, Germany

Publication Outlet

MIS Quarterly, Vol. 45 No. 3, pp. 1581-1602

Abstract

With the rise of machine learning (ML), humans are no longer the only ones capable of learning and contributing to an organization's stock of knowledge. We study how organizations can coordinate human learning and ML in order to learn effectively as a whole. Based on a series of agent-based simulations, we find that, first, ML can reduce an organization's demand for human explorative learning that is aimed at uncovering new ideas; second, adjustments to ML systems made by humans are largely beneficial, but this effect can diminish or even become harmful under certain conditions; and third, reliance on knowledge created by ML systems can facilitate organizational learning in turbulent environments, but this requires significant investments in the initial setup of these systems as well as adequately coordinating them with humans. These insights contribute to rethinking organizational learning in the presence of ML and can aid organizations in reallocating scarce resources to facilitate organizational learning in practice.

Keywords

Artificial Intelligence, Machine Learning, Human-Machine Coordination, Organizational Learning, Simulation, Agent-based Modeling

License

Table 12: Summary of results regarding organizational learning effectiveness

Research Questions	**Findings and Propositions**	**Implications**
RQ1: The Role of Human Exploration in the Presence of ML Systems	ML systems with a high initial learning capability reduce the need for human exploration (see P1).	• ML systems' ability to take over explorative tasks counters learning myopia, allowing humans to learn at their preferred pace. • Organizations should consider the reallocation of R&D resources to the initial setup of ML systems.
RQ2: Reconfiguration of ML Systems by Humans	Humans' learning behavior moderates the nonlinear effect of reconfiguration intensity on organizational learning effectiveness. For ML systems with a: • **Low initial learning capability:** If humans engage in exploitation (exploration), this effect is positive and decreases (increases) in strength with increasing reconfiguration intensity (see P2a). • **High initial learning capability:** If humans engage in exploitation, this effect decreases in strength with increasing reconfiguration intensity. If humans engage in exploration, the reconfiguration intensity has an inverted U-shaped effect (see P2b).	• Acquiring high levels of organizational knowledge requires at least a moderate amount of reconfiguration effort. • Humans should never be completely taken "out of the loop," even if tasks are largely automated. • As the deep problem understanding of domain experts is required for reconfiguration efforts, leaving reconfiguration of ML systems to the IT department alone is not sufficient.
RQ3: Coordinating Human Learning and ML Systems in Turbulent Environments	In turbulent environments, effective organizational learning with ML systems requires human exploration and a rapid codification of knowledgeable humans' beliefs. The more turbulent the environment, the more beneficial the rapid codification of beliefs offered by ML systems with a high initial learning capability will be (see P3).	• Reliance on knowledge created by ML systems can be beneficial for organizations in turbulent environments, reducing the need for more radical measures (e.g., forced personnel turnovers). • Significant investments in the initial setup of ML systems and appropriate coordination of humans and ML systems are required to materialize these beneficial effects.

7 Thesis Contributions and Conclusion

The overarching goal of this thesis was to improve the understanding of human-AI interaction. While research on AI-enabled systems has mainly focused on technical innovations in recent years (e.g., Ba et al., 2016; Sohn et al., 2020; Vaswani et al., 2017), the interaction between humans and AI-enabled systems has mostly been neglected so far. It is widely agreed among both IS and HCI researchers that the unique characteristics of AI-enabled systems, e.g., their probabilistic nature and predictive capabilities, require rethinking of existing guidelines for human-computer interaction (e.g., Maedche et al., 2019; Rai et al., 2019; Schuetz & Venkatesh, 2020; Yang et al., 2020). The four papers in this thesis aim to tackle this research gap from two perspectives: the impact of increasing human-AI interaction on an individual and organizational level. This section first provides a summary of the contributions of these papers from a theoretical and practical perspective. This thesis concludes with a vision of what human-AI interaction could look like further down the road and necessary research to achieve this vision.

7.1 Theoretical Contributions

The first part of this thesis (i.e., papers A and B) investigates the impact of increasing human-AI interaction on individuals. Paper A contributes to the growing literature on explainable AI and transparency of AI-enabled systems. While technical literature has come up with techniques for explaining the behavior of AI-enabled systems (e.g., Diakopoulos, 2016; Lundberg & Lee, 2017; Olah et al., 2018), it remains unclear mainly how end-users will perceive explanations in practice. Paper A provides two main contributions to XAI literature by (1) examining whether consumers exhibit a meaningful willingness-to-pay towards explainability features for AI-enabled systems and (2) developing a theoretical model for explaining the mechanisms behind the purchase decision. By conducting an online experiment in the context of credit applications, it is found that the majority of study participants would be willing to pay a median amount of 20€ for the explainability features offered by the fictional credit scoring provider. While the exact amount will differ between contexts, it can be expected that the general finding is transferable to other contexts where predictions from AI-enabled systems support assessing the personal attributes of individuals (e.g., their suitability for a given job or attributes influencing insurance premiums). Our theoretical model, based on the Theory of Planned Behavior (Ajzen, 1991), shows that increased user trust due to increased perceived transparency is the main driver for a positive evaluation of the AI-enabled system.

Paper B transfers the large body of research regarding the influence of technology on individual performance to the context of AI-enabled systems. The main theoretical contribution of this paper is a theoretical model based on the task-technology fit (TTF) theory (Goodhue & Thompson, 1995). The model is proposed based on the findings from expert interviews and considers the unique characteristics of AI-enabled systems, especially their reliance on statistical patterns identified in data to perform predictive tasks. In detail, the two constructs *data characteristics* and *task-data fit* are added to the classical TTF constructs to account for this dependence. Moreover, AI-specific dimensions are added to all theory constructs, including the classical TTF constructs *technology characteristics* and *task-technology fit*, to enable applications of the model in empirical research. While the focus was on explaining impacts on individual performance, the model is expected to be applicable for the group- or even organizational-level analyses of performance impacts related to deployments of AI-enabled systems.

The second part of this thesis (i.e., papers C and D) focuses on the organizational-level impact of increasing human-AI interaction. Paper C extends the fairness and organizational justice literature in that it examines the effects of deploying an AI-enabled system for recruiting purposes. While a lot of technical literature exists regarding how to ensure the fairness of AI-enabled systems (e.g., d'Alessandro et al., 2017; Mehrabi et al., 2021; Olteanu et al., 2019), it remains unclear whether AI-enabled will mitigate or reinforce existing biases in organizational decision-making (e.g., Maedche et al., 2019; Rai et al., 2019; Raisch & Krakowski, 2020). The study is based on an online experiment in which the candidate rankings of recruiters supported by a CV recommender system and unsupported recruiters are compared. It is shown that the introduction of the CV recommender system leads to a more consistent candidate selection which tends to include candidates with a higher level of objectively measurable skills (e.g., relevant working experience). Since objectivity and consistency are two primary components of procedural justice (e.g., Gilliland, 1993; Leventhal, 1980), the study proposes that AI-enabled systems can help ensure procedural justice in organizational decision-making.

Paper D examines the effects of AI-enabled systems on organizational learning. Previous studies have not reached a consensus on whether IT systems facilitate or hinder organizational learning, especially the creation and dissemination of knowledge (e.g., Alavi & Leidner, 2001; Robey et al., 2000; Schultze & Leidner, 2002). Moreover, these studies neglect the unique characteristics of AI-enabled systems, especially their ability to learn and contribute knowledge (e.g., Ransbotham et al., 2020; Seidel et al., 2019). Based on the results of a series of agent-based simulations, the primary theoretical contributions of paper D are derived in the form of three propositions. First, it is proposed that AI-enabled systems can take over explorative tasks, which reduces the need for human exploration and thus counters learning myopia (i.e., human's tendency to favor exploitation over exploration). Second, it is proposed that the reconfiguration of AI-enabled systems constitutes an essential activity for organizations. Third, it is suggested that the knowledge created by AI-enabled systems is beneficial for organizations in turbulent environments, reducing the need for alternative measures to deal with turbulence such as forced personnel turnover.

7.2 Practical Contributions

Apart from these theoretical contributions, the results derived from papers contained in this thesis also have significant implications for practitioners. The first part of this thesis (i.e., papers A and B) informs practitioners about how to improve the interaction between individuals and AI-enabled systems. Paper A shows that a significant willingness to pay exists for explainability features of AI-enabled systems. The findings can serve as a starting point for providers of these systems (e.g., credit scoring providers) to evaluate whether to deploy explainability features. Here, the study reveals two potential benefits of including these features. First, the experiment results suggest that the features can constitute an additional revenue stream, e.g., in the form of premium features. Moreover, the willingness to pay identified in the study can serve as an anchor for price determination of these features by decision-makers such as product managers. Second, the study provides evidence that an increase in perceived transparency leads to increased user trust and a more favorable attitude towards the AI-enabled system. Thus, deploying explainability features might positively influence essential business metrics such as user retention and engagement, as suggested by multiple studies (e.g., Chiu et al., 2012; Gounaris, 2005; Han & Hyun, 2015).

Paper B proposes a theoretical model for determining the task-technology fit as the primary determinant for utilization and performance impacts. While the model's primary purpose is to enable empirical research studies, practitioners can also apply it to evaluate possible use cases for deploying AI-enabled systems. One way of applying the model could be to evaluate use cases concerning the identified dimensions of each model construct to identify potential roadblocks. For example, developers of an AI-enabled system for supporting medical diagnostics might find ensuring the representability of collected datasets, particularly challenging. At the same time, comprehensibility might be an essential requirement for use cases that underlie strict regulatory requirements (e.g., finance). In addition, quantifying the dimensions with some scale enables practitioners to compare potential use cases for AI-enabled systems. All in all, our model can be valuable in the early stages of project management, which usually comprise feasibility evaluation and return-on-investment estimation (e.g., Amershi, Begel, et al., 2019; Bernardi et al., 2019).

In the second part of this thesis (i.e., papers C and D), the organizational-level impacts of increasing human-AI interaction are examined. From the results of these studies, practitioners can draw insights about how best to deploy AI-enabled systems to achieve organizational-level goals. Paper C suggests that organizations should consider deploying CV recommender systems for selecting candidates in their recruiting processes. The experimental results show that using these systems might serve three purposes. First, the objectivity of the candidate selection process might be increased, leading to a selection of applicants with higher levels of relevant skills (e.g., working experience). This way, organizations can prevent hiring the wrong candidates due to biases in the selection process, reducing costs for the organization. Second, candidate selection might become more consistent across different recruiters, reducing commonly occurring biases in the selection process (e.g., gender biases). Third, through partial automation of early phases in the selection process, recruiters might be enabled to focus on more cognitively challenging tasks such as in-person interviews. This would also reduce the burden on organizations challenged by the increasing amount of available applicant data (e.g., Black & van Esch, 2020; van Esch et al., 2019). While the online experiment in paper C is limited to the context of candidate selection, the results can be expected to be transferable to similar use cases in human resource management, e.g., the identification of candidates for promotions.

Paper D conducted several agent-based simulations to determine the optimal course for organizations to facilitate human-AI interaction concerning the critical activity of accumulating knowledge (i.e., organizational learning). From these simulations, two main insights can be drawn for practitioners. First, it was shown that AI-enabled systems could take over explorative learning tasks. This finding suggests that organizations should reconsider their resource allocations towards R&D projects. Shifting explorative responsibilities towards AI-enabled systems (e.g., discovering potential drug candidates) requires that organizations invest more heavily in developing these systems (e.g., by acquiring vast amounts of high-quality data and ML expertise). Second, it was found that the reconfiguration of deployed AI-enabled systems represents a key challenge for organizations. For organizations, this entails that both developers and domain experts should always be kept "in the loop" to improve the performance of AI-enabled systems continually. Example activities conducted by humans include identifying new data requirements, tuning the learning algorithms, and monitoring model performance in production (Amershi, Begel, et al., 2019).

7.3 Conclusion

Artificial intelligence is becoming increasingly prevalent in our private and professional lives. Technological innovations have enabled AI-enabled systems to take over tasks previously reserved for humans. However, how the interaction between humans and AI-enabled systems should be designed remains an understudied issue. This thesis, consisting of four qualitative and quantitative studies, aims to address this research gap by examining the impact of increasing human-AI interaction on an individual and organizational level. Thus, it constitutes an essential contribution to human-AI interaction research, which sits at the intersection of disciplines such as information systems, human-computer interaction, and computer science. Future research can build on this thesis' results. While some future research suggestions are offered in each of the included papers, the following paragraphs introduce some further ideas for upcoming research, both conceptual and empirical.

On a conceptual level, more research is needed to understand which aspects most strongly influence the interaction between humans and AI-enabled systems. This thesis addressed two focus points each regarding the individual-level (transparency and performance) and organizational-level (organizational justice and organizational learning) impacts of increasing human-AI interaction. However, it remains unclear how these aspects can be integrated into a holistic theory of human-AI interaction and which other aspects need to be investigated in future research. While some early work exists to describe human-AI interaction (e.g., Amershi, Weld, et al., 2019; Rzepka & Berger, 2018; Yang et al., 2020), the proposed frameworks and guidelines are not yet empirically validated and do not include perspectives from all relevant research fields. Multidisciplinary research will be essential to identify potential drivers and outcomes of human-AI interaction.

While conceptual research will be the key to deriving a holistic picture of what drivers and outcomes human-AI interaction encompasses, empirical research will be needed to derive principles and theories for how to best design human-AI interaction in practice. Here, the design science research methodology will play an essential role in coming up with design principles for different aspects of human-AI interaction. The developed design principles can be validated in different contexts in close collaboration with practitioners, primarily via field and lab experiments (Karahanna et al., 2018). These research efforts should be primarily conducted in high-stakes contexts such as healthcare and finance, which can hugely benefit from deploying AI-enabled systems and come with the highest risk. Early empirical research on human-AI interaction exists (e.g., Fügener et al., 2021; Grønsund & Aanestad, 2020; Sturm, Koppe, et al., 2021; Yang et al., 2019), but more will be needed to maximize the potential of this groundbreaking technology.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Adams, J. S. (1965). *Inequity In Social Exchange* (pp. 267–299). https://doi.org/10.1016/S0065-2601(08)60108-2

Adkins, C. L., Russell, C. J., & Werbel, J. D. (1994). Judgments of fit in the selection process: The role of work value congruence. *Personnel Psychology*, *47*(3), 605–623. https://doi.org/10.1111/j.1744-6570.1994.tb01740.x

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, *17*(6), 734–749. https://doi.org/10.1109/TKDE.2005.99

Ågerfalk, P. J. (2020). Artificial intelligence as digital agency. *European Journal of Information Systems*, *29*(1), 1–8. https://doi.org/10.1080/0960085X.2020.1721947

Aggarwal, C. C. (2016). *Recommender Systems*. Springer International Publishing. https://doi.org/10.1007/978-3-319-29659-3

Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). What to Expect from Artificial Intelligence Technology. *MIT Sloan Management Review*, *58*(3), 22–26. https://doi.org/10.7551/mitpress/11645.003.0008

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Ala'raj, M., & Abbod, M. F. (2016). A new hybrid ensemble credit scoring model based on classifiers consensus system approach. *Expert Systems with Applications*, *64*, 36–55. https://doi.org/10.1016/j.eswa.2016.07.017

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, *25*(1), 107–136. https://doi.org/10.2307/3250961

Almalis, N. D., Tsihrintzis, G. A., Karagiannis, N., & Strati, A. D. (2015). FoDRA — A new content-based job recommendation algorithm for job seeking and recruiting. *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, 1–7. https://doi.org/10.1109/IISA.2015.7388018

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software engineering for machine learning: A case study. *Proceedings of the 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice*, 291–300. https://doi.org/10.1109/ICSE-SEIP.2019.00042

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for Human-AI Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3290605.3300233

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989. https://doi.org/10.1177/2F1461444816676645

Andrews, W., Sau, M., Dekate, C., Mullen, A., Brant, K. F., Revang, M., & Plummer, D. C. (2018). *Predicts 2018: Artificial Intelligence*. Gartner. https://www.gartner.com/en/documents/3827163

Arnold, V., & Sutton, S. G. (1998). Arnold, V., & Sutton, S. G. (1998). The theory of technology dominance: Understanding the impact of intelligent decision aids on decision maker's judgments. *Advances in Accounting Behavioral Research*, *1*(3), 175–194.

Arvey, R. D., & Renz, G. L. (1992). Fairness in the selection of employees. *Journal of Business Ethics*, *11*(5–6), 331–340. https://doi.org/10.1007/BF00870545

Åslund, O., & Skans, O. N. (2012). Do Anonymous Job Application Procedures Level the Playing Field? *ILR Review*, *65*(1), 82–107. https://doi.org/10.1177/001979391206500105

Ba, J., Kiros, J. R., & Hinton, G. E. (2016). Layer Normalization. *ArXiv.Org*, *abs/1607.06450*.

Bamberger, P. A. (2018). AMD—Clarifying what we are about and where we are going. *Academy of Management Discoveries*, *4*(1), 1–10. https://doi.org/10.5465/amd.2018.0003

Bansal, S., Srivastava, A., & Arora, A. (2017). Topic Modeling Driven Content Based Jobs Recommendation Engine for Recruitment Industry. *Procedia Computer Science*, *122*, 865–872. https://doi.org/10.1016/j.procs.2017.11.448

Barredo Arrieta, A., Díaz-Rodríguez, N., del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Baylor, D., Breck, E., Cheng, H.-T., Fiedel, N., Foo, C. Y., Haque, Z., Haykal, S., Ispir, M., Jain, V., Koc, L., Koo, C. Y., Lew, L., Mewald, C., Modi, A. N., Polyzotis, N., Ramesh, S., Roy, S., Whang, S. E., Wicke, M., … Zinkevich, M. (2017). TFX: A TensorFlow-based production-scale machine learning platform. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1387–1395. https://doi.org/10.1145/3097983.3098021

Bean, R. (2019). *Demystifying artificial intelligence in the corporation*. Forbes. https://www.forbes.com/sites/ciocentral/2019/10/09/demystifying-artificial-intelligence-ai-in-the-corporation-forbes/#76f168df6016

Bernardi, L., Mavridis, T., & Estevez, P. (2019a). 150 Successful Machine Learning Models. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1743–1751. https://doi.org/10.1145/3292500.3330744

Bernardi, L., Mavridis, T., & Estevez, P. (2019b). 150 Successful Machine Learning Models. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1743–1751. https://doi.org/10.1145/3292500.3330744

Berner, C., Brockman, G., Chan, B., Cheung, V., Dębiak, P., Dennison, C., Farhi, D., Fischer, Q., Hashme, S., Hesse, C., Józefowicz, R., Gray, S., Olsson, C., Pachocki, J., Petrov, M., Pinto, H. P. d. O., Raiman, J., Salimans, T., Schlatter, J., … Zhang, S. (2019). *Dota 2 with Large Scale Deep Reinforcement Learning*.

Bhatt, G. D., & Zaveri, J. (2002). The enabling role of decision support systems in organizational learning. *Decision Support Systems*, *32*(3), 297–309. https://doi.org/10.1016/S0167-9236(01)00120-8

Bhatt, U., Xiang, A., Sharma, S., Weller, A., Taly, A., Jia, Y., Ghosh, J., Puri, R., Moura, J. M. F., & Eckersley, P. (2020). Explainable machine learning in deployment.

Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 648–657. https://doi.org/10.1145/3351095.3375624

Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, *84*, 317–331. https://doi.org/10.1016/j.patcog.2018.07.023

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer Science and Business Media.

Black, J. S., & van Esch, P. (2020). AI-enabled recruiting: What is it and how should a manager use it? *Business Horizons*, *63*(2), 215–226. https://doi.org/10.1016/j.bushor.2019.12.001

Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, *46*, 109–132. https://doi.org/10.1016/j.knosys.2013.03.012

Branch. (2019). *The power of financial access*. https://branch.co/about

Breidert, C., Hahsler, M., & Reutterer, T. (2006). A review of methods for measuring willingness-to-pay. *Innovative Marketing*, *2*(4), 8–32. https://pdfs.semanticscholar.org/6645/b7b08ec530f1201211dee550d045d3318e3d.pdf

Bretz, R. D., Ash, R. A., & Dreher, G. F. (1989). Do people make the place? An examination of the attraction-selection-attrition hypothesis. *Personnel Psychology*, *42*(3), 561–581. https://doi.org/10.1111/j.1744-6570.1989.tb00669.x

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*(6370), 1530–1534. https://doi.org/10.1126/science.aap8062

Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, *12*(4), 331–370. https://doi.org/10.1023/A:1021240730564

Burrell, J. (2015). How the machine “thinks”: Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 1–12. https://doi.org/10.2139/ssrn.2660674

Caldwell, D. F., & O’Reilly, C. A. (1990). Measuring person-job fit with a profile-comparison process. *Journal of Applied Psychology*, *75*(6), 648–657. https://doi.org/10.1037/0021-9010.75.6.648

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin Company.

Carter, S., Armstrong, Z., Schubert, L., Johnson, I., & Olah, C. (2019). Activation atlas. *Distill*, *4*(3), e15. https://doi.org/10.23915/distill.00015

Chai, Y., & Li, W. (2019). Towards deep learning interpretability: a topic modeling approach. *Proceedings of 40th International Conference on Information Systems*.

Chakraborti, T., Kulkarni, A., Sreedharan, S., Smith, D. E., & Kambhampati, S. (2019). Explicability? legibility? predictability? transparency? privacy? security? the emerging landscape of interpretable agent behavior. *Proceedings of the International Conference on Automated Planning and Scheduling*, *29*, 86–96.

Chang, H. H. (2008). Intelligent agent’s technology characteristics applied to online auctions’ task: A combined model of TTF and TAM. *Technovation*, *28*(9), 564–577. https://doi.org/10.1016/j.technovation.2008.03.006

Chatman, J. A. (1989). Matching People and Organizations: Selection and Socialization in Public Accounting Firms. *Academy of Management Proceedings*, *1989*(1), 199–203. https://doi.org/10.5465/ambpp.1989.4980837

Chen, C., Lin, K., Rudin, C., Shaposhnik, Y., Wang, S., & Wang, T. (2018). An interpretable model with globally consistent explanations for credit risk. *NIPS 2018 Workshop on Challenges and Opportunities for AI in Financial Services: The Impact of Fairness, Explainability, Accuracy, and Privacy*, 1–10.

Cheng, H. F., Wang, R., Zhang, Z., O’Connell, F., Gray, T., Harper, F. M., & Zhu, H. (2019). Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290605.3300789

Chernev, A. (2003). Reverse pricing and online price elicitation strategies in consumer choice. *Journal of Consumer Psychology*, *13*(1), 51–62. https://doi.org/10.1207/s15327663jcp13-1&2_05

Chiu, C. M., Hsu, M. H., Lai, H., & Chang, C. M. (2012). Re-examining the influence of trust on online repeat purchase intention: The moderating role of habit and its antecedents. *Decision Support Systems*, *53*(4), 835–845. https://doi.org/10.1016/j.dss.2012.05.021

Christin, A. (2017). Algorithms in practice: Comparing web journalism and criminal justice. *Big Data & Society*, *4*(2), 205395171771885. https://doi.org/10.1177/2053951717718855

Cohen, R. L. (1987). Distributive justice: Theory and research. *Social Justice Research*, *1*(1), 19–40. https://doi.org/10.1007/BF01049382

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, *86*(3), 386–400. https://doi.org/10.1037/0021-9010.86.3.386

Crowston, K., & Bolici, F. (2019). Impacts of machine learning on work. *Proceedings of the 52nd Hawai'i International Conference on System Sciences*, 5961–5970.

d'Alessandro, B., O'Neil, C., & LaGatta, T. (2017). Conscientious Classification: A Data Scientist's Guide to Discrimination-Aware Classification. *Big Data*, *5*(2), 120–134. https://doi.org/10.1089/big.2016.0048

David, P., Alan, M., & Randy, G. (1998). *Computational intelligence: A logical approach*. Oxford University Press.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management Science*, *35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

de Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., van den Driessche, G., Lakshminarayanan, B., Meyer, C., Mackinder, F., Bouton, S., Ayoub, K., Chopra, R., King, D., Karthikesalingam, A., … Ronneberger, O. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, *24*(9), 1342–1350. https://doi.org/10.1038/s41591-018-0107-6

Dean, A., Voss, D., & Draguljić, D. (2017). *Design and Analysis of Experiments*. Springer International Publishing. https://doi.org/10.1007/978-3-319-52250-0

Dennis, A. R., Wixom, B. H., & Vandenberg, R. J. (2001). Understanding fit and appropriation effects in group support systems via meta-analysis. *MIS Quarterly*, *25*(2), 167–193. https://doi.org/10.2307/3250928

Denrell, J., & March, J. G. (2001). Adaptation as Information Restriction: The Hot Stove Effect. *Organization Science*, *12*(5), 523–538. https://doi.org/10.1287/orsc.12.5.523.10092

Derrow-Pinion, A., She, J., Wong, D., Lange, O., Hester, T., Perez, L., Nunkesser, M., Lee, S., Guo, X., Wiltshire, B., Battaglia, P. W., Gupta, V., Li, A., Xu, Z., Sanchez-Gonzalez, A., Li, Y., & Velickovic, P. (2021). ETA Prediction with Graph Neural Networks in Google Maps. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 3767–3776. https://doi.org/10.1145/3459637.3481916

Deutsch, M. (1975). Equity, Equality, and Need: What Determines Which Value Will Be Used as the Basis of Distributive Justice? *Journal of Social Issues*, *31*(3), 137–149. https://doi.org/10.1111/j.1540-4560.1975.tb01000.x

Dhaliwal, J. S., & Benbasat, I. (1996). The use and effects of knowledge-based system explanations: theoretical foundations and a framework for empirical evaluation. *Information Systems Research*, *7*(3), 342–363. https://doi.org/10.1287/isre.7.3.342

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, *59*(2), 56–62. https://doi.org/10.1145/2844110

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science*, *64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Dodgson, M., Gann, D. M., & Phillips, N. (2013). Organizational learning and the technology of foolishness: The case of virtual worlds at IBM. *Organization Science*, *24*(5), 1358–1376. https://doi.org/10.1287/orsc.1120.0807

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *ArXiv.Org*, *stat.ML*, 1–13.

Eckhardt, A., Laumer, S., Maier, C., & Weitzel, T. (2014). The transformation of people, processes, and IT in e-recruiting. *Employee Relations*, *36*(4), 415–431. https://doi.org/10.1108/ER-07-2013-0079

Edwards, J. R. (1991). Person-job fit: A conceptual integration, literature review, and methodological critique. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (6th ed., pp. 283–357). John Wiley & Sons.

Elliot, B., & Andrews, W. (2017). *A framework for applying AI in the enterprise*. Gartner. https://www.gartner.com/ngw/globalassets/en/information-technology/documents/insights/a-framework-for-applying-ai-in-the-enterprise.pdf

Elofson, G. S., & Konsynski, B. R. (1993). Performing organizational learning with machine apprentices. *Decision Support Systems*, *10*(2), 109–119. https://doi.org/10.1016/0167-9236(93)90033-Y

Eurostat. (2018). *Internet access and use statistics - households and individuals*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Internet_access_and_use_statistics_-_households_and_individuals

Fang, C., Lee, J., & Schilling, M. A. (2010). Balancing exploration and exploitation through structural design: The isolation of subgroups and organizational learning. *Organization Science*, *21*(3), 625–642. https://doi.org/10.1287/orsc.1090.0468

Färber, F., Weitzel, T., & Keim, T. (2003). An Automated Recommendation Approach toSelection in Personnel Recruitment. *AMCIS 2003 Proceedings*, 302.

Fedyk, A. (2016). *How to Tell If Machine Learning Can Solve Your Business Problem*. Harvard Business Review. https://hbr.org/2016/11/how-to-tell-if-machine-learning-can-solve-your-business-problem

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly*, *48*(1), 94–118. https://doi.org/10.2307/3556620

Fernandez, C., & Provost, F. (2019). Counterfactual explanations for data-driven decisions. *Proceedings of 40th International Conference on Information Systems*.

FICO. (2019a). *Explainable machine learning challenge*. https://community.fico.com/s/explainable-machine-learning-challenge

FICO. (2019b). *FICO® Score X Data*. https://www.fico.com/en/products/fico-score-x-data

Fleming, N. (2018). How artificial intelligence is changing drug discovery. *Nature*, *557*(7707), S55–S57. https://doi.org/10.1038/d41586-018-05267-x

Forbes Insights. (2018). *On your marks: Business leaders prepare for arms race in artificial intelligence*. Forbes. https://www.forbes.com/sites/insights-intelai/2019/05/22/welcome-from-forbes-to-a-special-exploration-of-ai-issue-6/#67864c574650

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation. *Information Systems Research*. https://doi.org/10.1287/isre.2021.1079

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. https://doi.org/10.1016/j.jbusres.2015.12.008

Fuller, R. M., & Dennis, A. R. (2009). Does fit matter? The impact of task-technology fit and appropriation on team performance in repeated tasks. *Information Systems Research*, *20*(1), 2–17. https://doi.org/10.1287/isre.1070.0167

Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *Proceedings of The 33rd International Conference on Machine Learning*, *48*, 1050–1059.

Gay, L. R., Mill, G. E., & Airasian, P. W. (2012). *Educational research: competencies for analysis and applications*. Pearson Education, Inc.

Gebauer, J., & Ginsburg, M. (2009). Exploring the black box of task-technology fit. *Communications of the ACM*, *52*(1), 130–135.

Gebauer, J., Shaw, M. J., & Gribbins, M. L. (2010). Task-technology fit for mobile information systems. *Journal of Information Technology*, *25*(3), 259–272. https://doi.org/10.1057/jit.2010.10

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., III, H. D., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, *64*(12), 86–92. https://doi.org/10.1145/3458723

Gedikli, F., Jannach, D., & Ge, M. (2014). How should I explain? A comparison of different explanation types for recommender systems. *Journal of Human-Computer Studies*, *72*(4), 367–382. https://doi.org/10.1016/j.ijhcs.2013.12.007

George, J. F. (2004). The theory of planned behavior and internet purchasing. *Internet Research*, *14*(3), 198–212. https://doi.org/10.1108/10662240410542634

Ghasemaghaei, M., Hassanein, K., & Turel, O. (2017). Increasing firm agility through the use of data analytics: The role of fit. *Decision Support Systems*, *101*, 95–105. https://doi.org/10.1016/j.dss.2017.06.004

Ghorbani, A., Wexler, J., Zou, J. Y., & Kim, B. (2019). Towards Automatic Concept-based Explanations. *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS)*, *32*.

Gilliland, S. W. (1993). The Perceived Fairness of Selection Systems: An Organizational Justice Perspective. *The Academy of Management Review*, *18*(4), 694. https://doi.org/10.2307/258595

Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. *Journal of Computational and Graphical Statistics*, *24*(1), 44–65. https://doi.org/10.1080/10618600.2014.907095

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). *Explaining and Harnessing Adversarial Examples*. https://arxiv.org/abs/1412.6572

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, *19*(2), 213–233. https://doi.org/10.2307/249689

Goodman, B., & Flaxman, S. (2017). European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation." *AI Magazine*, *38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Google. (2019). *Perspectives on issues in AI governance*. https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf

Gounaris, S. P. (2005). Trust and commitment influences on customer retention: insights from business-to-business services. *Journal of Business Research*, *58*(2), 126–140. https://doi.org/10.1016/S0148-2963(03)00122-X

Green, B., & Chen, Y. (2019). Disparate Interactions. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 90–99. https://doi.org/10.1145/3287560.3287563

Greenberg, J., & Colquitt, J. A. (2005). *Handbook of Organizational Justice*. Psychology Press.

Gregor, S. (2001). Explanations from knowledge-based systems and cooperative problem solving: an empirical study. *International Journal of Human-Computer Studies*, *54*(1), 81–105. https://doi.org/10.1006/ijhc.2000.0432

Gregor, S., & Benbasat, I. (1999). Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS Quarterly*, *23*(4), 497–530. https://doi.org/10.2307/249487

Grgić-Hlača, N., Engel, C., & Gummadi, K. P. (2019). Human Decision Making with Machine Assistance. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–25. https://doi.org/10.1145/3359280

Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic Information Systems*, *29*(2), 101614. https://doi.org/10.1016/j.jsis.2020.101614

Grossklags, J., & Acquisti, A. (2007). When 25 Cents is too much: An experiment on Willingness-To-Sell and Willingness-To-Protect personal information. *Proceedings of 6th Workshop Econom. Inform. Secury (WEIS '07).*

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A survey of methods for explaining black box models. *ACM Computing Surveys*, *51*(5), 93. https://doi.org/10.1145/3236009

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On Calibration of Modern Neural Networks. *Proceedings of the 34th International Conference on Machine Learning*, *70*, 1321–1330.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Prentice Hall.

Han, H., & Hyun, S. S. (2015). Customer retention in the medical tourism industry: Impact of quality, satisfaction, trust, and price reasonableness. *Tourism Management*, *46*, 20–29. https://doi.org/10.1016/j.tourman.2014.06.003

Hann, I.-H., & Terwiesch, C. (2003). Measuring the frictional costs of online transactions: the case of a name-your-own-price channel. *Management Science*, *49*(11), 1563–1579. https://doi.org/10.1287/mnsc.49.11.1563.20586

Hansen, T., Jensen, J. M., & Solgaard, H. S. (2004). Predicting online grocery buying intention: a comparison of the theory of reasoned action and the theory of planned behavior. *International Journal of Information Management*, *24*(6), 539–550. https://doi.org/10.1016/j.ijinfomgt.2004.08.004

Hatch, N. W., & Dyer, J. H. (2004). Human capital and learning as a source of sustainable competitive advantage. *Strategic Management Journal*, *25*(12), 1155–1178. https://doi.org/10.1002/smj.421

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*(4), 408–420. https://doi.org/10.1080/03637750903310360

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Heess, N., Sriram, S., Lemmon, J., Merel, J., Wayne, G., Tassa, Y., Erez, T., Wang, Z., Eslami, S. M., Riedmiller, M., & Silver, D. (2017). Emergence of locomotion behaviours in rich environments. *ArXiv.Org*, *preprint a*, 1–14.

Hohman, F., Head, A., Caruana, R., DeLine, R., & Drucker, S. M. (2019). Gamut: A design probe to understand how data scientists understand machine learning models. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290605.3300809

Höök, K. (2000). Steps to take before intelligent user interfaces become real. *Interacting with Computers*, *12*(4), 409–426. https://doi.org/10.1016/S0953-5438(99)00006-5

Horvitz, E. (1999). Principles of mixed-initiative user interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems the CHI Is the Limit - CHI '99*, 159–166. https://doi.org/10.1145/302979.303030

Howard, J., & Ruder, S. (2018). *Universal Language Model Fine-tuning for Text Classification*.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, *49*(1), 30–50. https://doi.org/10.1007/s11747-020-00749-9

Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ArXiv.Org*, *abs/1502.03167*.

ITU. (2018). *United Nations activities on artificial intelligence (AI)*. https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-UNACT-2018-1-PDF-E.pdf

Iyengar, K., Sweeney, J. R., & Montealegre, R. (2015). Information Technology Use as a Learning Mechanism: The Impact of IT Use on Knowledge Transfer Effectiveness, Absorptive Capacity, and Franchisee Performance. *MIS Quarterly*, *39*(3), 615–641. https://doi.org/10.25300/MISQ/2015/39.3.05

Judge, T. A., & Cable, D. M. (1997). Applicant personality, organizational culture, and organizational attraction. *Personnel Psychology*, *50*(2), 359–394. https://doi.org/10.1111/j.1744-6570.1997.tb00912.x

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Junglas, I., Abraham, C., & Watson, R. T. (2008). Task-technology fit for mobile locatable information systems. *Decision Support Systems*, *45*(4), 1046–1057.

Kane, G. C., & Alavi, M. (2007). Information technology and organizational learning: An investigation of exploration and exploitation processes. *Organization Science*, *18*(5), 796–812. https://doi.org/10.1287/orsc.1070.0286

Kappelman, L., L. Johnson, V., Maurer, C., Guerra, K., McLean, E., Torres, R., Snyder, M., & Kim, K. (2020). The 2019 SIM IT Issues and Trends Study. *MIS Quarterly Executive*, *19*(1), 69–104. https://doi.org/10.17705/2msqe.00026

Karahanna, E., Benbasat, I., Bapna, R., & Rai, A. (2018). Opportunities and challenges for different types of online experiments. *MIS Quarterly*, *42*(4), iii–xi.

Karimi, J., Somers, T. M., & Gupta, Y. P. (2004). Impact of environmental uncertainty and task characteristics on user satisfaction with data. *Information Systems Research*, *15*(2), 175–193. https://doi.org/10.1287/isre.1040.0022

Kendall, M. G. (1946). *The advanced theory of statistics*. Charles Griffin and Co.

Kim, B., Khanna, R., & Koyejo, O. (2016). Examples are not enough, learn to criticize! Criticism for interpretability. *Proceedings of the 30th Conference on Neural Information Processing Systems (NeurIPS)*.

Kim, B., Wattenberg, M., Gilmer, J., Cai, C., Wexler, J., Viegas, F., & Sayres, R. (2018). Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (TCAV). *Proceedings of the 35th International Conference on Machine Learning*, 2668–2677.

Kirk, R. E. (2012). *Experimental Design*. SAGE Publications.

Kleinberg, J. M., Mullainathan, S., & Raghavan, M. (2017). Inherent Trade-Offs in the Fair Determination of Risk Scores. *ArXiv.Org*, *abs/1609.05807*.

Koh, P. W., & Liang, P. (2017). Understanding black-box predictions via influence functions. *Proceedings of the 34th International Conference on Machine Learning, ICML 2017*, *4*, 2976–2987.

Koufaris, M., & Hampton-Sosa, W. (2004). The development of initial trust in an online company by new customers. *Information & Management*, *41*(3), 377–397. https://doi.org/10.1016/j.im.2003.08.004

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. v., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, *13*, 8–17. https://doi.org/10.1016/j.csbj.2014.11.005

Krasnova, H., Hildebrand, T., & Guenther, O. (2009). Investigating the value of privacy on online social networks: conjoint analysis. *Proceedings of the 30th International Conference on Information Systems*, 173.

Kreditech. (2018). *Through a unique data scoring algorithm based on machine learning*. https://www.kreditech.com/

Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology*, *49*(1), 1–49. https://doi.org/10.1111/j.1744-6570.1996.tb01790.x

Kristof-Brown, A. L. (2000). Perceived applicant fit: Distinguishing between recruiters' perceptions of person-job and person-organization fit. *Personnel Psychology*, *53*(3), 643–671. https://doi.org/10.1111/j.1744-6570.2000.tb00217.x

Kristof-Brown, A. L., Zimmermann, R. D., & Johnson, E. C. (2005). Consequences of individuals' fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology*, *58*(2), 281–342. https://doi.org/10.1111/j.1744-6570.2005.00672.x

Krizhevsky, A., Sutskever, I., & Geoffrey E., H. (2012). Imagenet classification with deep convolutional neural networks. *Proceedings of Advances in Neural Information Processing Systems*, 1097–1105. https://doi.org/10.1109/5.726791

Kruppa, J., Schwarz, A., Arminger, G., & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. *Expert Systems with Applications*, *40*(13), 5125–5131. https://doi.org/10.1016/j.eswa.2013.03.019

Kruse, L., Wunderlich, N., & Beck, R. (2019). Artificial intelligence for the financial services industry: What challenges organizations to succeed. *Proceedings of the 52nd Hawai'i International Conference on System Sciences*, 6408–6417.

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. Penguin.

Kwon, B. C., Choi, M. J., Kim, J. T., Choi, E., Kim, Y. bin, Kwon, S., Sun, J., & Choo, J. (2019). RetainVis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. *IEEE Transactions on Visualization and Computer Graphics*, *25*(1), 299–309. https://doi.org/10.1109/TVCG.2018.2865027

Lavie, D., Stettner, U., & Tushman, M. L. (2010). Exploration and exploitation within and across organizations. *The Academy of Management Annals*, *4*(1), 109–155. https://doi.org/10.1080/19416521003691287

Lee, C.-C., Cheng, H. K., & Cheng, H.-H. (2007). An empirical study of mobile commerce in insurance industry: Task-technology fit and individual differences. *Decision Support Systems*, *43*(1), 95–110.

Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. *Minds & Machines*, *17*(4), 391–444. https://doi.org/10.1007/s11023-007-9079-x

Leventhal, G. S. (1980). What Should Be Done with Equity Theory? In *Social Exchange* (pp. 27–55). Springer US. https://doi.org/10.1007/978-1-4613-3087-5_2

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*, *14*(S2), 95–112. https://doi.org/10.1002/smj.4250141009

Levitt, B., & March, J. G. (1988). Organizational learning. *Annual Review of Sociology*, *14*(1), 319–338. https://doi.org/10.1146/annurev.so.14.080188.001535

Li, F.-C., Wang, P.-K., & Liu, Y. C. (2009). Credit scoring based on hybrid data mining classification. *Proceedings of the 15th Americas Conference on Information Systems*, 181.

Li, L., & Zhang, P. (2005). The Intellectual Development of Human-Computer Interaction Research: A Critical Assessment of the MIS Literature (1990-2002). *Journal of the Association for Information Systems*, *6*(11), 227–292. https://doi.org/10.17705/1jais.00070

Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing Design Practices for Explainable AI User Experiences. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3313831.3376590

Lipton, Z. C. (2016). The mythos of model interpretability. *ArXiv.Org*, *cs.LG*, 1–9.

Liu, J., Dai, R., Wei, X. (David), & Li, Y. (2016). Information revelation and customer decision-making process of repeat-bidding name-your-own-price auction. *Decision Support Systems*, *90*, 46–55. https://doi.org/10.1016/j.dss.2016.06.018

Logg, J. M. (2019). *Using Algorithms to Understand the Biases in Your Organization*. Harvard Business Review. https://hbr.org/2019/08/using-algorithms-to-understand-the-biases-in-your-organization

Lops, P., de Gemmis, M., & Semeraro, G. (2011). Content-based Recommender Systems: State of the Art and Trends. In *Recommender Systems Handbook* (pp. 73–105). Springer US. https://doi.org/10.1007/978-0-387-85820-3_3

Lu, Y., el Helou, S., & Gillet, D. (2013). A recommender system for job seeking and recruiting website. *Proceedings of the 22nd International Conference on World Wide Web - WWW '13 Companion*, 963–966. https://doi.org/10.1145/2487788.2488092

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Proceedings of Advances in Neural Information Processing Systems*, 4768–4777.

Lyytinen, K., Nickerson, J. v, & King, J. L. (2021). Metahuman systems = humans + machines that learn. *Journal of Information Technology*, *36*(4), 427–445. https://doi.org/10.1177/0268396220915917

Maddox, W. J., Garipov, T., Izmailov, P., Vetrov, D., & Wilson, A. G. (2019). A Simple Baseline for Bayesian Uncertainty in Deep Learning. *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS)*, 13153–13164.

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: an integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277–301. https://doi.org/10.1080/14639220500337708

Madsen, A. (2019). Visualizing memorization in RNNs. *Distill*, *4*(3), e16. https://doi.org/10.23915/distill.00016

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., & Söllner, M. (2019). AI-Based Digital Assistants. *Business & Information Systems Engineering*, *61*(4), 535–544. https://doi.org/10.1007/s12599-019-00600-8

Malinowski, J., Keim, T., Wendt, O., & Weitzel, T. (2006). Matching People and Jobs: A Bilateral Recommendation Approach. *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, 137c–137c. https://doi.org/10.1109/HICSS.2006.266

Mao, J.-Y., & Benbasat, I. (2000). The use of explanations in knowledge-based systems: Cognitive perspectives and a process-tracing analysis. *Journal of Management Information Systems*, *17*(2), 153–179. https://doi.org/10.1080/07421222.2000.11045646

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*(1), 71–87. https://doi.org/10.1287/orsc.2.1.71

Martens, D., Baesens, B., van Gestel, T., & Vanthienen, J. (2007). Comprehensible credit scoring models using rule extraction from support vector machines. *European Journal of Operational Research*, *183*(3), 1466–1476. https://doi.org/10.1016/j.ejor.2006.04.051

Martens, D., & Provost, F. (2014). Explaining data-driven document classifications. *MIS Quarterly*, *38*(1), 73–99. https://doi.org/10.1088/1751-8113/44/8/085201

Maruping, L. M., & Agarwal, R. (2004). Managing team interpersonal processes through technology: A Task-technology fit perspective. *Journal of Applied Psychology*, *89*(6), 975–990. https://doi.org/10.1037/0021-9010.89.6.975

McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. S., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F. J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C. J., King, D., … Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, *577*(7788), 89–94. https://doi.org/10.1038/s41586-019-1799-6

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, *54*(6), 1–35. https://doi.org/10.1145/3457607

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*.

Miller, K. D., Zhao, M., & Calantone, R. J. (2006). Adding interpersonal learning and tacit knowledge to March's exploration–exploitation model. *Academy of Management Journal*, *49*(4), 709–722. https://doi.org/10.5465/amj.2006.22083027

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability and Transparency*, 220–229. https://doi.org/10.1145/3287560.3287596

Mitchell, T. (1997). *Machine learning*. McGraw-Hill. https://doi.org/10.1145/242224.242229

Mothilal, R. K., Sharma, A., & Tan, C. (2020). Explaining machine learning classifiers through diverse counterfactual explanations. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 607–617. https://doi.org/10.1145/3351095.3372850

Muchinsky, P. M., & Monahan, C. J. (1987). What is person-environment congruence? Supplementary versus complementary models of fit. *Journal of Vocational Behavior*, *31*(3), 268–277. https://doi.org/10.1016/0001-8791(87)90043-1

Murray, A., Rhymer, J., & Sirmon, D. G. (2021). Humans and Technology: Forms of Conjoined Agency in Organizations. *Academy of Management Review*, *46*(3), 552–571. https://doi.org/10.5465/amr.2019.0186

Nilsson, N. J. (1998). *Artificial intelligence: A new synthesis*. Morgan Kaufmann. https://doi.org/10.1016/C2009-0-27773-7

Norman, D. A. (1994). How might people interact with agents. *Communications of the ACM*, *37*(7), 68–71. https://doi.org/10.1145/176789.176796

Ochmann, J., & Laumer, S. (2019). Fairness as a Determinant of AI Adoption in Recruiting: An Interview-based Study. *DIGIT 2019 Proceedings*, e16.

Olah, C., Satyanarayan, A., Johnson, I., Carter, S., Schubert, L., Ye, K., & Mordvintsev, A. (2018). The building blocks of interpretability. *Distill*, *3*(3), e10. https://doi.org/10.23915/distill.00010

Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *Frontiers in Big Data*, *2*. https://doi.org/10.3389/fdata.2019.00013

Ostroff, C. (1993). The Effects of Climate and Personal Influences on Individual Behavior and Attitudes in Organizations. *Organizational Behavior and Human Decision Processes*, *56*(1), 56–90. https://doi.org/10.1006/obhd.1993.1045

Ötting, S. K., & Maier, G. W. (2018). The importance of procedural justice in Human–Machine Interactions: Intelligent systems as new decision agents in organizations. *Computers in Human Behavior*, *89*, 27–39. https://doi.org/10.1016/j.chb.2018.07.022

Parkes, A. (2013). The effect of task-individual-technology fit on user attitude and performance: An experimental investigation. *Decision Support Systems*, *54*(2), 997–1009. https://doi.org/10.1016/j.dss.2012.10.025

Pazzani, M. J., & Billsus, D. (n.d.). Content-Based Recommendation Systems. In *The Adaptive Web* (pp. 325–341). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-72079-9_10

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. https://doi.org/10.1037/0021-9010.88.5.879

Pourghasemi, H. R., Kariminejad, N., Amiri, M., Edalat, M., Zarafshar, M., Blaschke, T., & Cerda, A. (2020). Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. *Scientific Reports*, *10*(1), 3203. https://doi.org/10.1038/s41598-020-60191-3

Rai, A., Constantinides, P., & Sarker, S. (2019). Editor's comments: Next-generation digital platforms: Toward human–AI hybrids. *MIS Quarterly*, *43*(1), iii–ix.

Raisch, S., Birkinshaw, J., Probst, G., & Tushman, M. L. (2009). Organizational ambidexterity: Balancing exploitation and exploration for sustained performance. *Organization Science*, *20*(4), 685–695. https://doi.org/10.1287/orsc.1090.0428

Raisch, S., & Krakowski, S. (2020). Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review*, *forthcoming*. https://doi.org/10.5465/2018.0072

Ramezani, M., Bergman, L. D., Thompson, R., Burke, R. D., & Mobasher, B. (2008). Selecting and applying recommendation technology. *International Workshop on Recommendation and Collaboration in Conjunction with 2008 International ACM Conference on Intelligent User Interfaces*, 613–620.

Ransbotham, S., Khodabandeh, S., Kiron, D., Candelon, F., Chu, M., & Lafountain, B. (2020). *Expanding AI's impact with organizational learning*. MIT Sloan Management Review. https://sloanreview.mit.edu/projects/expanding-ais-impact-with-organizational-learning/

Reyes, M., Meier, R., Pereira, S., Silva, C. A., Dahlweid, F.-M., Tengg-Kobligk, H. von, Summers, R. M., & Wiest, R. (2020). On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities. *Radiology: Artificial Intelligence*, *2*(3), e190043. https://doi.org/10.1148/ryai.2020190043

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you?: Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778

Ribeiro, M. T., Singh, S., & Guestrin, C. (2018). Anchors: High-precision model-agnostic explanations. *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, 1527–1535.

Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (2011). *Recommender Systems Handbook*. Springer US. https://doi.org/10.1007/978-0-387-85820-3

Robey, D., Boudreau, M. C., & Rose, G. M. (2000). Information technology and organizational learning: A review and assessment of research. *Accounting, Management and Information Technologies*, *10*(2), 125–155. https://doi.org/10.1016/S0959-8022(99)00017-X

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*(10), 1276–1284. https://doi.org/10.1037/0003-066X.44.10.1276

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Russell, S., & Norvig, P. (2016). *Artificial intelligence: a modern approach* (3rd ed.). Addison-Wesley.

Rynes, S., & Gerhart, B. (1990). Interviewer assessments of applicant “fit”: An exploratory investigation. *Personnel Psychology*, *43*(1), 13–35. https://doi.org/10.1111/j.1744-6570.1990.tb02004.x

Rzepka, C., & Berger, B. (2018). User interaction with AI-enabled systems: a systematic review of IS research. *Proceedings of the 39th International Conference on Information Systems*.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

Satell, G. (2018). *How to Make an AI Project More Likely to succeed*. Harvard Business Review. https://hbr.org/2018/07/how-to-make-an-ai-project-more-likely-to-succeed

Schafer, J. ben, Frankowski, D., Herlocker, J., & Sen, S. (n.d.). Collaborative Filtering Recommender Systems. In *The Adaptive Web* (pp. 291–324). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-72079-9_9

Schmelzer, R. (2019). *The AI-Enabled Future*. Forbes. https://www.forbes.com/sites/cognitiveworld/2019/10/17/the-ai-enabled-future/#634cc1083339

Schneider, B. (1983). An Interactionist Perspective on Organizational Effectiveness. In *Organizational Effectiveness* (pp. 27–54). Elsevier. https://doi.org/10.1016/B978-0-12-157180-1.50007-0

Schneider, B. (1987). The people make the place. *Personnel Psychology*, *40*(3), 437–453. https://doi.org/10.1111/j.1744-6570.1987.tb00609.x

Schneider, B. (2001). Fits About Fit. *Applied Psychology*, *50*(1), 141–152. https://doi.org/10.1111/1464-0597.00051

Schreiner, M., & Hess, T. (2015). Why are consumers willing to pay for privacy? An application of the privacy-freemium model to media companies. *Proceedings of the 23rd European Conference on Information Systems*.

Schuetz, S., & Venkatesh, V. (2020). Research perspectives: The rise of human machines: How cognitive computing systems challenge assumptions of user-system interaction. *Journal of the Association for Information Systems*, *21*(2), 460–482. https://doi.org/10.17705/1jais.00608

Schultze, U., & Leidner, D. E. (2002). Studying knowledge management in information systems research: Discourses and theoretical assumptions. *MIS Quarterly*, *26*(3), 213–242. https://doi.org/10.2307/4132331

Schumm, W. R., Pratt, K. K., Hartenstein, J. L., Jenkins, B. A., & Johnson, G. A. (2013). Determining statistical significance (alpha) and reporting statistical trends: controversies, issues, and facts [1]. *Comprehensive Psychology*, *2*(1), Article 10. https://doi.org/10.2466/03.CP.2.10

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J. F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2503–2511.

Seidel, S., Berente, N., Lindberg, A., Lyytinen, K., & Nickerson, J. v. (2019). Autonomous tools and design: A triple-loop approach to human-machine learning. *Communications of the ACM*, *62*(1), 50–57. https://doi.org/10.1145/3210753

Sekiguchi, T. (2004). Person-organization fit and person-job fit in employee selection: A review of the literature. *Osaka Keidai Ronshu*, *54*(6), 179196.

Shook, J., Smith, R., & Antonio, A. (2018). Transparency and fairness in machine learning applications. *Texas A&M Journal of Property Law*, *4*, 443–462. https://doi.org/10.3868/s050-004-015-0003-8

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*(1), 60. https://doi.org/10.1186/s40537-019-0197-0

Sidorova, A., & Rafiee, D. (2019). AI agency risks and their mitigation through business process management: a conceptual framework. *Proceedings of the 52nd Hawai'i International Conference on System Sciences*, 5837–5845.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, *362*(6419), 1140–1144. https://doi.org/10.1126/science.aar6404

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, *550*(7676), 354–359. https://doi.org/10.1038/nature24270

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*(1), 99. https://doi.org/10.2307/1884852

Simon, H. A. (1960). *The new science of management decision.* Harper & Brothers. https://doi.org/10.1037/13978-000

Sohn, K., Berthelot, D., Li, C.-L., Zhang, Z., Carlini, N., Cubuk, E. D., Kurakin, A., Zhang, H., & Raffel, C. (2020). FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. *ArXiv.Org*, *abs/2001.07685*.

Spann, M., Skiera, B., & Schäfers, B. (2004). Measuring individual frictional costs and willingness-to-pay via name-your-own-price mechanisms. *Journal of Interactive Marketing*, *18*(4), 22–36. https://doi.org/10.1002/dir.20022

Strohmeier, S., & Piazza, F. (2015). *Artificial Intelligence Techniques in Human Resource Management—A Conceptual Exploration* (pp. 149–172). https://doi.org/10.1007/978-3-319-17906-3_7

Sturm, T., Gerlach, J., Pumplun, L., Mesbah, N., Peters, F., Tauchert, C., Nan, N., & Buxmann, P. (2021). Coordinating Human and Machine Learning for Effective Organization Learning. *MIS Quarterly*, *45*(3), 1581–1602. https://doi.org/10.25300/MISQ/2021/16543

Sturm, T., Koppe, T., Scholz, Y., & Buxmann, P. (2021). The Case of Human-Machine Trading as Bilateral Organizational Learning Learning. *Proceedings of the 42nd International Conference on Information Systems*. https://aisel.aisnet.org/icis2021/ai_business/ai_business/3

Suh, B., & Han, I. (2002). Effect of trust on customer acceptance of internet banking. *Electronic Commerce Research and Applications*, *1*(3–4), 247–263.

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. *Proceedings of the 34th International Conference on Machine Learning*, 3319–3328.

Suresh, H., & Guttag, J. v. (2019). A Framework for Understanding Unintended Consequences of Machine Learning. *ArXiv.Org*, *abs/1901.10002*.

Taylor, S., & Todd, P. A. (1995a). Understanding information technology usage: a test of competing models. *Information Systems Research*, *6*(2), 144–176. https://doi.org/10.1287/isre.6.2.144

Taylor, S., & Todd, P. A. (1995b). Understanding information technology usage: a test of competing models. *Information Systems Research*, *6*(2), 144–176. https://doi.org/10.1287/isre.6.2.144

The Economist. (2018). *For artificial intelligence to thrive, it must explain itself.* https://www.economist.com/science-and-technology/2018/02/15/for-artificial-intelligence-to-thrive-it-must-explain-itself

Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, *31*(2), 447–464. https://doi.org/10.1007/s12525-020-00441-4

Thielsch, M. T., Träumer, L., & Pytlik, L. (2012). E-recruiting and fairness: the applicant's point of view. *Information Technology and Management*, *13*(2), 59–67. https://doi.org/10.1007/s10799-012-0117-x

van Esch, P., Black, J. S., & Ferolie, J. (2019). Marketing AI recruitment: The next phase in job application and selection. *Computers in Human Behavior*, *90*, 215–222. https://doi.org/10.1016/j.chb.2018.09.009

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Proceedings of Advances in Neural Information Processing Systems*, 5998–6008.

Veale, M., van Kleek, M., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 440. https://doi.org/10.1145/3173574.3174014

Venkatesh, V., Morris, M., Gordon, B., & Davis, F. (2003a). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Venkatesh, V., Morris, M., Gordon, B., & Davis, F. (2003b). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Verma, S., & Rubin, J. (2018). Fairness Definitions Explained. *2018 IEEE/ACM International Workshop on Software Fairness (FairWare)*, 1–7. https://doi.org/10.23919/FAIRWARE.2018.8452913

Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., Choi, D. H., Powell, R., Ewalds, T., Georgiev, P., Oh, J., Horgan, D., Kroiss, M., Danihelka, I., Huang, A., Sifre, L., Cai, T., Agapiou, J. P., Jaderberg, M., … Silver, D. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, *575*(7782), 350–354. https://doi.org/10.1038/s41586-019-1724-z

von Krogh, G. (2018). Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. *Academy of Management Discoveries*, *4*(4), 404–409. https://doi.org/10.5465/amd.2018.0084

Wachter, S., Mittelstadt, B., & Russell, C. (2017). *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*. http://arxiv.org/abs/1711.00399

Wagner, A., Wessels, N., Buxmann, P., & Krasnova, H. (2018). Putting a price tag on personal information - a literature review. *Proceedings of the 51st Hawaii International Conference on System Sciences*, 3760–3769.

Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019). Designing theory-driven user-centric explainable AI. *Conference on Human Factors in Computing Systems - Proceedings*, 1–15. https://doi.org/10.1145/3290605.3300831

Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*, *48*, 144–156. https://doi.org/10.1016/j.jmsy.2018.01.003

Wang, W., & Benbasat, I. (2016). Empirical Assessment of Alternative Designs for Enhancing Different Types of Trusting Beliefs in Online Recommendation Agents. *Journal of Management Information Systems*, *33*(3), 744–775. https://doi.org/10.1080/07421222.2016.1243949

Wang, W., Qiu, L., Kim, D., & Benbasat, I. (2016). Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust. *Decision Support Systems*, *86*, 48–60. https://doi.org/10.1016/j.dss.2016.03.007

Wei, J., He, J., Chen, K., Zhou, Y., & Tang, Z. (2017). Collaborative filtering and deep learning based recommendation system for cold start items. *Expert Systems with Applications*, *69*, 29–39. https://doi.org/10.1016/j.eswa.2016.09.040

Wilk, S. L., & Sackett, P. R. (1996). Longitudinal analysis of ability-job complexity fit and job change. *Personnel Psychology*, *49*(4), 937–967. https://doi.org/10.1111/j.1744-6570.1996.tb02455.x

Wongpinunwatana, N., Ferguson, C., & Bowen, P. (2000). An experimental investigation of the effects of artificial intelligence systems on the training of novice auditors. *Managerial Auditing Journal*, *15*(6), 306–318.

Wu, J., & Liu, D. (2007). The effects of trust and enjoyment on intention to play online games. *Journal of Electronic Commerce Research*, *8*(2), 128–140.

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376301

Yang, Q., Steinfeld, A., & Zimmerman, J. (2019). Unremarkable AI: Fiting intelligent decision support into critical, clinical decision-making processes. *Conference on Human Factors in Computing Systems - Proceedings*, 1–11. https://doi.org/10.1145/3290605.3300468

Ye, R. L., & Johnson, P. E. (1995). The impact of explanation facilities on user acceptance of expert systems advice. *MIS Quarterly*, *19*(2), 157–172. https://doi.org/10.2307/249686

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, *37*(2), 197–206. https://doi.org/10.1086/651257

Zhou, J. (2017). Data mining for individual consumer credit default prediction under e-commence context: a comparative study. In *Proceedings of 38th International Conference on Information Systems* (p. 22).

Zhu, D., Prietula, M. J., & Hsu, W. L. (1997). When processes learn: Steps toward crafting an intelligent organization. *Information Systems Research*, *8*(3), 302–317. https://doi.org/10.1287/isre.8.3.302

Zigurs, I., & Buckland, B. K. (1998). A theory of task/technology fit and group support systems effectiveness. *MIS Quarterly*, *22*(3), 313–334. https://doi.org/10.2307/249668