Exploring Human and Artificial Intelligence Collaboration and Its Impact on Organizational Performance: A Multi-Level Analysis

TECHNISCHE UNIVERSITÄT DARMSTADT

Department of Law and Economics
at the Technical University of Darmstadt

Dissertation

by

Timo Sturm

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

First Assessor:	Prof. Dr. Peter Buxmann
Second Assessor:	Prof. Dr. Alexander Benlian
Darmstadt	2023

Sturm, Timo: Exploring Human and Artificial Intelligence Collaboration and its Impact on Organizational Performance: A Multi-Level Analysis
Darmstadt, Technische Universität Darmstadt,
Year thesis published in TUprints 2023
Date of the viva voce: 02/02/2023

Declaration of Authorship

I hereby declare that the submitted thesis is my own work. All quotes, whether word for 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.

Timo Sturm

Darmstadt, November 28, 2022

Abstract

To achieve great performance and ensure their long-term survival, organizations must successfully act in and adapt to the reality that surrounds them, which requires organizations to learn effectively. For decades, organizations have relied exclusively on human learning for this purpose. With today's rise of machine learning (ML) systems as a modern form of artificial intelligence (AI) and their ability to autonomously learn and act, ML systems can now also contribute to this vital process, offering organizations an alternative way to learn. Although organizations are increasingly adopting ML systems within a wide range of processes, we still know surprisingly little about how the learning of humans and ML systems affects each other and how their mutual learning affects organizational performance. Although a significant amount of research has addressed ML, existing research leaves it largely unclear whether and when humans and ML systems act as beneficial complementarities or as mutual impediments within the context of learning. This is problematic, as the (mis)use of ML systems may corrupt an organization's central process of learning and thus impair the organizational adaptation that is crucial for organizational survival.

To help organizations facilitate useful synergies of humans and ML systems, this dissertation explores humans' and ML systems' idiosyncrasies and their bilateral interplay. As research on organizational learning has demonstrated, the key to managing such dynamics is the effective coordination of the ones who learn. The studies that were conducted for this dissertation therefore aim to uncover virtuous and vicious dynamics between humans and ML systems and how these dynamics can be managed to increase organizational performance. To take a holistic perspective, this dissertation explores three central levels of analysis.

The first level of analysis deals with performance impacts on the *individual level*. Here, the analysis focuses on two essential issues. First, the availability of ML systems as an alternative to humans requires organizations to rethink their problem delegation strategies. Organizations can benefit the most from the relative strengths of humans and ML systems if they are able to delegate problems to those whose expertise and capabilities best fit the problem. This requires organizations to develop an understanding of the problem characteristics that point to problems that are better (or less) suited to being solved by ML systems than by humans. Using a qualitative interview approach, the first study identifies central criteria and procedural artifacts and synthesizes these into a framework for identifying and evaluating problems in ML contexts. The framework provides a theoretical basis to help inform research about delegation decisions between humans and ML systems by unpacking problem nuances that decisively render problems suitable for ML systems. Building on these insights, a subsequent qualitative analysis explores how the dependency between a human and an ML system with respect to the delegated problem affects performance outcomes. The theoretical model that is proposed explains individual performance gains that result from ML systems' use as a function of the fit between task, data, and technology characteristics. The model highlights how idiosyncrasies of an ML system can affect a human expert's task execution performance when the expert bases her/his task execution on the ML system's contributions. This study provides first empirical evidence on controllable levers for managing involved dependencies to increase individual performance.

The second level of analysis focuses on performance impacts on the *group level*. In contrast to traditional (non-ML) information systems, ML systems' unique learning ability enables them to contribute independently to team endeavors, joining groups as active members that can affect group dynamics through their own contributions. Thus, in a third study, a digital trace analysis is conducted to explore the dynamics of a real-world case in which a group of human traders and a productively trading reinforcement ML system collaborate during trading. The studied case reveals that bilateral learning between multiple humans and an ML system can increase trading performance, which appears to be the result of an emerging virtuous cycle between the humans and the ML system. The findings demonstrate that the interactions between the humans and the ML system can lead to group performance that outperforms the individual trading of either the humans or the ML system. However, in order to achieve this, organizations must effectively coordinate the knowledge transfer and the roles of the involved humans and the ML system.

The third level of analysis focuses on performance impacts on the *organization level*. As ML systems increasingly contribute to organizational processes in all areas of the organization, changes in the organization's fundamental concepts are likely to occur, and these may affect the organization's overall performance. In a fourth study, a series of agent-based simulations are therefore used to explore the dynamics of organization-wide interactions between humans and ML systems. The results imply that ML systems can help stimulate the pursuit of innovative directions, liberating humans from exploring unorthodox ideas. The results also show that the alignment of human learning and ML is largely beneficial but can, under certain conditions, become detrimental to organizations. The findings emphasize that effective coordination of humans and ML systems that takes environmental conditions into account can determine the positive and negative impacts of ML systems on organization-level performance.

The analyses included in this dissertation highlight that it is precisely the unique differences between humans and ML systems that often seem to make them better complements than substitutes for one another. The secret to unleashing the true potential of ML systems may therefore lie in effectively coordinating the differences between humans and ML systems within their bilateral relationship to produce virtuous cycles of mutual improvement. This dissertation is a first step toward developing theory and guidance on coordinating the dynamics between humans and ML systems, with the aim of helping to rethink collaboration theory in the era of AI.

Abstract (German Version)

Um hohe Leistungen zu erzielen und ihr langfristiges Überleben zu sichern, müssen Organisationen erfolgreich in der sie umgebenden Realität agieren und sich an diese anpassen, was ein effektives Lernen der Organisationen erfordert. Jahrzehntelang haben sich Organisationen dabei ausschließlich auf das Lernen von Menschen verlassen. Mit dem Aufkommen des maschinellen Lernens (ML) als moderne Form der künstlichen Intelligenz (KI) und ihrer Fähigkeit, autonom zu lernen und zu handeln, können ML-Systeme nun auch zu diesem wichtigen Prozess beitragen und Unternehmen eine alternative Möglichkeit des Lernens bieten. Obwohl Unternehmen zunehmend ML-Systeme in einer Vielzahl von Prozessen einsetzen, wissen wir nur erstaunlich wenig darüber, wie sich das Lernen von Menschen und ML-Systemen gegenseitig beeinflusst und wie ihr gemeinsames Lernen die Unternehmensleistung prägt. Obwohl es bereits viel Forschung zu ML gibt, bleibt weitgehend unklar, ob und wann Menschen und ML-Systeme als nützliche Ergänzungen oder als schädliche Hindernisse beim gemeinsamen Lernen wirken. Dies ist insofern problematisch, als der Einsatz von ML-Systemen den zentralen Prozess des Lernens beeinträchtigen und die für das Überleben von Organisationen entscheidende Anpassung erschweren kann.

Um Organisationen dabei zu helfen, nützliche Synergien zwischen Menschen und ML-Systemen zu ermöglichen, erforscht diese Dissertation die Eigenarten von Menschen und ML-Systemen und deren bilaterales Zusammenspiel. Wie die Forschung zum organisationalen Lernen gezeigt hat, liegt der Schlüssel zur Bewältigung derartiger Dynamiken in der effektiven Koordination der beteiligten Lernenden. Die durchgeführten Studien zielen daher darauf ab, vorteilhafte und schädliche Dynamiken von Menschen und ML-Systemen aufzudecken und zu untersuchen, wie diese Dynamiken koordiniert werden können, um die organisationale Leistung zu steigern. Um eine ganzheitliche Perspektive zu fördern, werden drei zentrale Analyseebenen erforscht.

Die erste Ebene der Analyse befasst sich mit Leistungsauswirkungen auf *individueller Ebene*. Hier konzentriert sich die Analyse auf zwei wesentliche Aspekte. Zunächst erfordert die Verfügbarkeit von ML-Systemen als Alternative zu Menschen, dass Organisationen ihre Delegationsstrategien überdenken. Unternehmen können am meisten von den Stärken von Menschen und ML-Systemen profitieren, wenn sie in der Lage sind, Probleme an diejenigen zu delegieren, deren Fachwissen und Fähigkeiten am geeignetsten für die Lösung des bestehenden Problems sind. Dies setzt voraus, dass Unternehmen ein Verständnis für Problemcharakteristika entwickeln, die auf Probleme hinweisen, die sich besser (oder schlechter) für die Lösung mittels ML-Systemen anstelle von Menschen eignen. Mithilfe einer qualitativen Interviewstudie werden zentrale Kriterien und Verfahrensartefakte identifiziert und zu einer Rahmenstruktur für die Identifizierung und Bewertung von Problemen in ML-Kontexten zusammengefasst. Die Rahmenstruktur bildet eine theoretische Grundlage für die Erforschung von Delegationsentscheidungen zwischen Menschen und ML-Systemen, indem sie die wesentlichen Merkmale von Problemen herausarbeitet, die diese für die Anwendung von ML-Systemen geeignet erscheinen lassen. Aufbauend auf diesen Erkenntnissen wird in einer weiteren qualitativen Analyse untersucht, wie sich die Abhängigkeit zwischen einem Menschen und einem ML-System bei delegierten Problemen auf die erzielte Leistung auswirkt. Es wird ein theoretisches Modell entwickelt, das individuelle Leistungssteigerungen aufgrund des Einsatzes von ML-Systemen als Funktion der Kompatibilität zwischen Aufgaben-, Daten- und Technologiemerkmalen erklärt. Das Modell verdeutlicht, wie die Eigenarten eines ML-Systems die Leistung eines menschlichen Experten bei der Aufgabenausführung beeinflussen können, wenn sie/er ihre/seine Ausführung auf die Beiträge des ML-Systems stützt. Die Studie liefert erste empirische Evidenz für kontrollierbare Einflussfaktoren, um zentrale Abhängigkeiten zu managen und die individuelle Leistung zu steigern.

Die zweite Ebene der Analyse konzentriert sich auf Leistungsauswirkungen auf *Gruppenebene*. Im Gegensatz zu konventionellen (Nicht-ML-)Informationssystemen können ML-Systeme aufgrund ihrer einzigartigen Lernfähigkeit eigenständig zu Teamvorhaben beitragen und sich Gruppen als aktive Mitglieder anschließen, die durch ihre eigenen Beiträge die Gruppendynamik beeinflussen können. Um die entstehende Dynamik zu erforschen, wird eine Digital-Trace-Analyse eines realen Falles durchgeführt, in welchem eine Gruppe von menschlichen Händlern und ein produktiv handelndes ML-System beim Wertpapierhandel zusammenarbeiten. Der untersuchte Fall zeigt, dass bilaterales Lernen zwischen mehreren Menschen und einem ML-System die Handelsleistung steigern kann, was aus einem sich entwickelnden positiven Kreislauf zwischen den Menschen und dem ML-System zu resultieren scheint. Die Ergebnisse zeigen, dass die Interaktionen zwischen den Menschen und dem ML-System zu einer Gruppenleistung führen können, die die individuelle Leistung der Menschen oder des ML-Systems übertrifft. Dies setzt jedoch voraus, dass Organisationen den Wissenstransfer und die Rollen der beteiligten Menschen und des ML-Systems effektiv koordinieren können.

Schließlich konzentriert sich die dritte Analyseebene auf Leistungsauswirkungen auf *Organisationsebene*. Da ML-Systeme in zunehmendem Maße zu organisationalen Prozessen in sämtlichen Bereichen einer Organisation beitragen, ist es wahrscheinlich, dass es zu Veränderungen in den grundlegenden Konzepten der Organisation kommt, was sich wiederum auf die Gesamtleistung der Organisation auswirken kann. Durch eine Reihe agentenbasierter Simulationen werden die Dynamiken der organisationsweiten Interaktionen zwischen Menschen und ML-Systemen erforscht. Die Ergebnisse deuten darauf hin, dass ML-Systeme dazu beitragen können, die Verfolgung innovativer Richtungen anzuregen und Menschen von der Erprobung unkonventioneller Ideen zu entlasten. Darüber hinaus wird gezeigt, dass die Verbindung von menschlichem Lernen und ML weitgehend vorteilhaft ist, aber unter bestimmten Bedingungen auch nachteilig für Organisationen sein kann. Die Ergebnisse zeigen, dass eine effektive Koordination von Menschen und ML-Systemen unter Berücksichtigung der Umweltbedingungen die positiven und negativen Auswirkungen von ML-Systemen auf die Leistung von Organisationen maßgeblich beeinflussen kann.

Die enthaltenen Analysen verdeutlichen, dass es gerade die einzigartigen Unterschiede zwischen Menschen und ML-Systemen sind, die dafür sorgen, dass sie einander oft besser ergänzen als ersetzen. Das Geheimnis zur Entfaltung des wahren Potenzials von ML-Systemen mag daher darin liegen, die Unterschiede zwischen Menschen und ML-Systemen innerhalb ihrer bilateralen Beziehung effektiv zu koordinieren, um positive Kreisläufe der gegenseitigen Verbesserung zu erzeugen. Diese Dissertation ist ein erster Schritt zur Entwicklung einer Theorie und Leitlinie für die Koordinierung der Dynamiken zwischen Menschen und ML-Systemen und soll dazu beitragen, Kollaborationstheorie im Zeitalter der KI neu zu überdenken.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AGI	Artificial general intelligence
AI	Artificial intelligence
AllGI	Allianz Global Investors
DSS	Decision support systems
E	(Model) Element
GDPR	General Data Protection Regulation
GSS	Group support systems
IS	Information systems
ISI	Institute for Scientific Information
IT	Information technology
ML	Machine learning
P	Proposition
RQ	Research question
R&D	Research and development
SJR	SCImago Journal Rank
TAM	Technology acceptance model
TTF	Task-technology fit
VHB-JQ3	VHB-JOURQUAL3

1 Introduction

Artificial intelligence (AI) is disrupting our economy and society, affecting all kinds of domains and industries (e.g., Benbya et al. 2021; Berente et al. 2021; Brynjolfsson and Mcafee 2016; Daugherty and Wilson 2018; Davenport and Kirby 2016; Ransbotham et al. 2020). However, although AI is being adopted in a wide variety of organizational processes, many organizations struggle to achieve the promised performance gains when they adopt AI, as they fail to exploit AI's great potential (e.g., Benbya et al., 2021; Berente et al., 2021; Fügener, Grahl, Gupta, & Ketter, 2021; Ransbotham et al., 2020; Teodorescu et al., 2021). To help advance our understanding of the impact of AI on organizational performance and to provide guidance on how to effectively manage AI, this dissertation explores the idiosyncrasies of AI, the influential role of the bilateral relationship that develops between humans and AI and gives rise to virtuous or vicious dynamics in human–AI collaborations, and the resulting managerial implications.

1.1 Overarching Motivation

To achieve great performance, organizations must be able to act and adapt effectively, which requires them to continuously learn about the reality that surrounds them (e.g., Argote et al., 2021; Levitt & March, 1988; March, 1991). This makes learning a crucial core process for organizations that forms organizational decisions, routines, and innovations—and thereby can even determine their long-term survival (e.g., Argote et al., 2021; Argote & Miron-Spektor, 2011; Huber, 1991; Levitt & March, 1988). However, organizations cannot learn on their own, but rely exclusively on the learning of their members (e.g., March, 1991; Simon, 1991). Thus, the key to enabling and sustaining worthwhile learning about their environment lies in organizations' facilitation and coordination of the learning and interactions of their members (e.g., Argote et al., 2021; Fang et al., 2010; Levitt & March, 1988; March, 1991, 2010; K. D. Miller & Lin, 2010). Unfortunately, the complex nature of reality makes learning about its causal structure a very challenging endeavor (e.g., Levinthal & March, 1993; March, 2010): Reality involves a mind-boggling and ambiguous number of variables, variable connections, and random variations (e.g., Benbya et al., 2020; March, 1994, 2010), which creates a level of complexity that often appears to lie beyond reach (e.g., Levinthal & March, 1993; March, 2010; G. A. Miller, 1956; Simon, 1972). The unraveling of reality's complex dynamics has therefore already become the nemesis of humanity's greatest minds (e.g., March, 2006, 2010; Simon, 1972). The limits of human cognition, which only allow observing and analyzing a strongly limited number of the variables, variable connections, and random variations of reality, constitute one of the greatest obstacles to learning about reality, and thereby inhibit the effectiveness of the decisions, routines, and crafted innovations that occur (e.g., Levinthal & March, 1993; March, 2010; March et al., 1991; G. A. Miller, 1956; Simon, 1972).

To overcome human limitations, an astounding number of technologies with increased analytical capabilities have been created (e.g., Berente et al., 2021; Lindebaum et al., 2020;

March, 2006). Presently, this endeavor has led to a particular form of modern AI known as *machine learning* (ML). Information systems (ISs) based on ML can *learn on their own* by deriving patterns from data to create models of reality that can be used to guide future behavior (Brynjolfsson & Mitchell, 2017; Mitchell, 1997; Russell & Norvig, 2021). ML systems have recently surpassed human intelligence in a variety of tasks, including Go (e.g., Silver et al., 2017), object detection in images (e.g., He et al., 2015), and complex multiplayer online games (e.g., Vinyals et al., 2019). ML systems are therefore often praised as a universal panacea to overcome the limits of human cognition (e.g., Benbya et al., 2021; Lindebaum et al., 2020). Attracted by this great potential, today's organizations are increasingly adopting ML systems (e.g., Brynjolfsson & Mitchell, 2017; Ransbotham et al., 2020). However, some recent disasters have made it apparent that ML systems may not be a universal panacea after all (e.g., Dolata et al., 2022; Fu et al., 2022; Kordzadeh & Ghasemaghaei, 2022; Marjanovic et al., 2022; Teodorescu et al., 2021). Indeed, ML systems come with their own strengths and weaknesses (as discussed in the studies included in this dissertation), which appear to render ML systems not necessarily a better but rather a *different* form of learners than humans (e.g., Balasubramanian et al., 2022; Berente et al., 2021; Ransbotham et al., 2020).

Due to the wide-ranging adoption of ML systems, humans and ML systems already learn side by side in organizations to jointly help them learn about their environment in order to act and adapt effectively (e.g., Brynjolfsson & Mitchell, 2017; Murray et al., 2021; Ransbotham et al., 2020; Seidel et al., 2019), and this requires organizations to effectively coordinate the collective learning of interacting humans and ML systems (e.g., Fügener, Grahl, Gupta, & Ketter, 2021; Ransbotham et al., 2020; Seidel et al., 2019; Sturm, Gerlach, et al., 2021). To date, however, it remains unclear whether and when humans' and ML systems' differences may be complementary or detrimental to one another (e.g., Benbya et al., 2021; Berente et al., 2021; Fügener, Grahl, Gupta, & Ketter, 2021; Schuetz & Venkatesh, 2020). In the best case, the reciprocal interplay of humans and ML systems can generate helpful synergies to overcome their respective shortcomings (e.g., Benbya et al., 2021; Schuetz & Venkatesh, 2020; Sturm, Gerlach, et al., 2021). In the worst case, however, they may corrupt each other and thus possibly become detrimental to the organizational learning in its entirety, thereby threatening organizations' performance and long-term survival (e.g., Balasubramanian et al., 2022; Fügener, Grahl, Gupta, & Ketter, 2021; Sturm, Gerlach, et al., 2021).

It is only recently that researchers have begun to examine how ML systems affect human behavior and vice versa (e.g., Grønsund & Aanestad, 2020; Lebovitz et al., 2021; Schuetz & Venkatesh, 2020; Seidel et al., 2019; Teodorescu et al., 2021). Research on the role of learning within collaborations between humans and ML systems still remains scarce, and the scholars involved in such research strongly emphasize the need for further analyses (e.g., Afiouni-Monla, 2019; Argote et al., 2021; Balasubramanian et al., 2022; Lyytinen et al., 2021; Ransbotham et al., 2020; Seidel et al., 2019; Sturm, Gerlach, et al., 2021; Sturm, Koppe, et al., 2021). As organizations may not only miss out on ML systems' potential but may even risk impeding the learning processes that are vital to their organizational behavior and survival, our lack of understanding on how to effectively coordinate the dynamics of humans and ML systems can have far-reaching consequences (e.g., Balasubramanian et al., 2022; Berente et al., 2021; Lyytinen et al., 2021; Schuetz & Venkatesh, 2020; Sturm, Gerlach, et al., 2021). To help uncover virtuous and vicious dynamics between humans and ML systems and to help organizations effectively manage these dynamics, I pursue the following overarching research objective: *In order to increase organizational performance, how can organizations effectively coordinate the learning of their human members and ML systems?*

1.2 Research Questions

As widespread collaboration between humans and ML systems can be expected to affect different parts of an organization, I further nuance the overarching research objective with the following research questions (RQs) to capture different levels of analysis. I adopt three levels of analysis that are commonly used to analyze changes in organizational behavior to help provide a holistic perspective on the phenomenon of collaboration between humans and ML systems in organizations: the (1) *individual*, (2) *group*, and (3) *organization* levels.

When organizations introduce ML systems, the *individual* level is affected first and foremost. As ML systems offer a different form of learning, organizations may use them as an alternative to human learning. Organizations should reap the most benefit from the different strengths of humans and ML systems if they are able to delegate learning problems to those whose expertise best fits the problem, which requires organizations to rethink their delegation strategies (e.g., Brynjolfsson & Mitchell, 2017; Fügener, Grahl, Gupta, & Keter, 2021; Lyytinen et al., 2021). To do so, organizations must develop an understanding of the problem characteristics that can help discern the better (or inferior) suitability of humans and ML systems for certain learning problems in order to guide their delegation decisions (e.g., Brynjolfsson & Mitchell, 2017; Jordan & Mitchell, 2015). To date, it remains unclear how organizations can distinguish among learning problems to effectively guide their delegation of these problems to humans or ML systems (e.g., Benbya et al., 2021; Berente et al., 2021; Fügener, Grahl, Gupta, & Keter, 2021; Lyytinen et al., 2021). To help advance our understanding of how to delegate problems to humans or ML systems, I ask the following RQ:

RQ1.1: *In order to increase individual performance, how can organizations effectively delegate learning problems between a human and an ML system?*

However, delegation between humans and ML systems is only half of the story of coordination on the individual level. The assumption of an "all-or-nothing" scenario neglects potential synergies that can emerge through the combination of humans and ML systems. If done correctly, collaboration can be more than just the sum of individual contributions by virtue of the additional value that can emerge from individuals' interactions (e.g., Argote et al., 2021; Lyles & Fiol, 1985; March, 1991). Here, research on task-technology fit (TTF) offers a theoretical lens that is a good fit for studying the interaction of a human and an ML system when a human bases her/his task execution on the outcomes of an ML system's actions. TTF theory highlights that the better a technology fits an individual's task, the more likely the technology is to increase the individual's performance (Goodhue & Thompson, 1995). While research on TTF has been applied to a wide variety of technological contexts (e.g., group support and mobile ISs; Gebauer et al., 2010; Zigurs & Buckland, 1998), ML systems are as yet unexplored. Prior findings on similar technologies (e.g., expert systems and data analytics; Ghasemaghaei et al., 2017; Wongpinunwatana et al., 2000) exist, but they are ill-suited to inform studies on ML systems because they do not capture the idiosyncrasies of ML systems.[1] To investigate the conditions that facilitate potential improvements of a human's performance through interaction with an ML system, I ask the following RQ:

RQ1.2: *In order to increase individual performance, how can organizations enable effective collaboration between a human and an ML system on a shared task?*

[1] As outlined in the theoretical background and further detailed in paper B.

To understand the consequences of organizational change on the *group* level, it does not suffice to extrapolate from separate individual perspectives (e.g., Argote & Miron-Spektor, 2011; Lyles & Fiol, 1985; March, 1991). Group-level analysis concerns a set of individuals working together to reach a mutual objective, creating value beyond the sum of the separate individuals' actions (e.g., Lyles & Fiol, 1985; March, 1991). Here, the intertwined (either mutually nurturing or hindering) actions of multiple individuals often create novel dynamics that cannot be easily anticipated from the individual perspective (e.g., Benbya et al., 2020; March, 1991). Group-level dynamics have been analyzed in a wide variety of contexts (e.g., Sarker & Valacich, 2010; Zigurs & Buckland, 1998), including diverse usages of different ISs (e.g., group support systems, corporate repositories; Alavi & Leidner, 2001; Kane & Alavi, 2007). To date, however, ISs have been viewed exclusively as a tool that only supports the learning within human groups (e.g., knowledge repositories that facilitate the accumulation and sharing of group members' knowledge; Kane & Alavi, 2007). As ML systems are capable of contributing their own learning to group endeavors, ML systems may join groups as additional active group members (e.g., Seeber et al., 2020). As a result, new group formations may emerge, potentially shifting, eliminating, and creating tasks and roles within groups (e.g., Fügener, Grahl, Gupta, & Ketter, 2021; Schuetz & Venkatesh, 2020; Seeber et al., 2020). Yet empirical evidence that allows unpacking such group dynamics between humans and ML systems remains scarce (e.g., Fügener, Grahl, Gupta, & Ketter, 2021). To help uncover virtuous and vicious group dynamics between humans and ML systems, I ask the following RQ:

RQ2: *In order to increase group performance, how can organizations effectively coordinate the group dynamics of humans and ML systems*?

Organization-level research focuses on fundamental concepts, such as the evolution of organizational norms, cultures, and strategies, that can have wide-ranging consequences for the organization as a whole. A strong cumulative tradition of organization-level research exists (e.g., how ISs can support organization-wide knowledge exchange and how digitalization strategies can be reliably aligned with organizational visions; Argote et al., 2021; Kohli & Melville, 2019). However, insights on the organization-level dynamics of ML systems and their consequences have not yet been established (e.g., Berente et al., 2021; Murray et al., 2021). As ML systems increasingly join humans in core learning processes and contribute what they have learned throughout the organization, changes in the organization's collective knowledge are likely to emerge, potentially affecting organization-wide concepts (e.g., Lyytinen et al., 2021; Murray et al., 2021; Teodorescu et al., 2021). To understand when and how such wide-ranging changes emerge from the learning of humans and ML systems within an organization, an analysis of organization-wide dynamics can inform us about related consequences and potential countermeasures (e.g., Kane & Alavi, 2007; March, 1991). To help understand the organization-wide role of learning dynamics between humans and ML systems, I ask the following RQ:

RQ3: *In order to increase organization-wide performance, how can organizations effectively coordinate humans and ML systems throughout the entire organization?*

1.3 Structure of this Dissertation

Aiming to help answer the RQs introduced above, this dissertation includes four research papers that have been published in peer-reviewed outlets, including three publications in conference proceedings and one journal article. In this section, the contributions and research approaches of the four papers are summarized and integrated into the overall structure of this dissertation.

Table 1 lists the four papers, referred to as A, B, C, and D, that are included in this dissertation. Paper A addresses RQ1.1, regarding individual performance, by exploring relevant procedural artifacts and key factors that can inform the delegation of problems between humans and ML systems, and then condenses these into a proposed framework that allows structured ideation and evaluation of problems that are suited to be solved using ML systems. Paper B, which also focuses on individual performance, tackles RQ1.2. This paper translates the widely renowned TTF theory to a context in which a human bases her/his task execution on an ML system's actions and thereby develops a nuanced version of TTF to capture the relevant idiosyncrasies of ML systems. The proposed theoretical model explains individual performance gains via use of ML systems as a function of the fit between task, data, and technology characteristics, thus contributing to a better understanding of the potential levers for generating synergies between a human and an ML system. Paper C, which focuses on group-level performance, aims to answer RQ2. Through a study of an autonomously trading ML system that joins a team of human traders, this paper uncovers several virtuous and vicious dynamics between the human traders and the ML system. The observed dynamics uncover coordination designs that increase the group-level trading performance. Finally, paper D focuses on organization-level performance and aims to answer RQ3 by exploring organization-wide dynamics between humans and ML systems and their effects on organization-level performance. Different coordination setups are observed over the lifetime of a simulated organization in which humans and ML systems learn side by side, jointly affecting the organization's overall stock of knowledge (i.e., its procedures, rules, forms, and norms; March 1991). The derived insights are contrasted in different environmental contexts to eventually inform organizational strategies for effective collaboration designs between humans and ML systems that will likely increase organization-wide performance.

Table 1: List of publications included in this dissertation

Paper A: Individual-level performance	Sturm, Timo, Fecho, Mariska, & Buxmann, Peter. (2021). **To use or not to use artificial intelligence? A framework for the ideation and evaluation of problems to be solved with artificial intelligence**. In *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS)*, Virtual Conference. VHB-JQ3[2] Ranking: C
Paper B: Individual-level performance	Sturm, Timo, & Peters, Felix. (2020). **The impact of artificial intelligence on individual performance: Exploring the fit between task, data, and technology**. In *Proceedings of the 28th European Conference on Information Systems (ECIS)*, a Virtual AIS Conference. VHB-JQ3 Ranking: B
Paper C: Group-level performance	Sturm, Timo, Koppe, Timo, Scholz, Yven, & Buxmann, Peter. (2021). **The case of human-machine trading as bilateral organizational learning**. In *Proceedings of the 42nd International Conference on Information Systems (ICIS)*, Austin, TX, USA. VHB-JQ3 Ranking: A
Paper D: Organization-level performance	Sturm, Timo, Gerlach, Jin P., Pumplun, Luisa, Mesbah, Neda, Peters, Felix, Tauchert, Christoph, Nan, Ning, & Buxmann, Peter. (2021). **Coordinating human and machine learning for effective organizational learning**. *Management Information Systems Quarterly* (*MIS Quarterly*), *45*(3), 1581–1602. https://doi.org/10.25300/MISQ/2021/16543 VHB-JQ3 Ranking: A+ (SJR: Q1 | Impact factor (2021): 8.553)

[2] The VHB-JOURQUAL3 (VHB-JQ3) ranking was selected by the Technical University of Darmstadt as the preferred source for assessing the quality of research papers in my doctoral study program. The VHB-JQ3 was published by the German Academic Association of Business Research in 2015, and this remained the latest VHB-JQ ranking at the time of writing this dissertation. For the journal publication, I also include two internationally recognized rankings to increase transparency: the SCImago Journal Rank (SJR) and the Institute for Scientific Information (ISI) impact factor.

Table 2 provides a detailed overview of the four papers' levels of analysis, research approaches, and theoretical backgrounds. As empirical evidence on the use of ML systems in organizations remains scarce, a variety of explorative research methods are used within the four papers, with the aim of contributing theoretical foundations that can inspire future research endeavors: First, *qualitative content analysis* is used in papers A and B. Through open-ended questions, qualitative content analysis allows experts to freely share their experiences and opinions, which are then codified in order to nuance existing theory for novel contexts. In both essays, the content analysis is based on a series of interviews with experts who are frequently involved in ML initiatives. The two studies contribute to developing a fundamental understanding of the use of ML systems and its impact on individual performance by condensing the experts' experiences across a multitude of industries and use cases. The uncovered impacts of ML systems' perils and pitfalls on a human's performance were then also used to inspire papers C and D. Second, *digital trace analysis* is used in paper C. Digital trace analysis follows an abductive reasoning approach that combines multiple iterations of quantitative and qualitative analyses. For the quantitative data, the actual trading behaviors of both human traders and an autonomous ML system were collected, analyzed, and then contrasted to gradually derive working hypotheses. To nuance and validate these working hypotheses, we then reached out to the traders and presented them with the derived insights, aiming to collect collateral qualitative data to explain the quantified trading behavior. Insights from the qualitative analyses then stimulated additional working hypotheses and quantitative analyses. This cycle of quantitative and qualitative analyses was performed until sufficiently justified hypotheses emerged. This approach allowed us to uncover the dynamics between the human traders and the ML system and provide initial explanations for their emergence. Third, a series of *agent-based simulations* were used in paper D. Relying on agent-based simulation allows the pursuit of three objectives: (1) to model a potential future state of an organization in which humans and ML systems learn equally side by side throughout the organization, (2) to observe the evolution of processes over an organization's complete lifetime (an observation that is largely obstructed in real-world empirical settings due to constraints in measuring capability and time), and (3) to adopt a holistic organization-wide perspective on emergent dynamics and derivable propositions.

Table 2: Outline of research papers

Paper	Level of Analysis	Research Approach	Theoretical Base
Paper A	Individual	Qualitative content analysis	Problem solving
Paper B			Task-technology fit (TTF)
Paper C	Group	Digital trace analysis	Organizational learning
Paper D	Organization	Agent-based simulation	

In addition to the four papers that are included in this dissertation, I co-authored the following publications in the fields of intelligent transportation, health care, and computer science:[3]

[3] Although these three publications are related to the overarching topic, they are not included in this dissertation because their essential focus is different from the focus of this dissertation (i.e., exploring the impacts of human–AI collaboration on organizational performance).

- Sturm, T., Krupitzer, C., Segata, M., & Becker, C. (2021). A taxonomy of optimization factors for platooning. *IEEE Transactions on Intelligent Transportation Systems*, *22*(10), 6097–6114. https://doi.org/10.1109/tits.2020.2994537
(VHB-JQ3: – | SJR: Q1 | Impact factor (2021): 9.551)
- Gawlitza, J., Sturm, T., Spohrer, K., Henzler, T., Akin, I., Schönberg, S., Borggrefe, M., Haubenreisser, H., & Trinkmann, F. (2019). Predicting pulmonary function testing from quantified computed tomography using machine learning algorithms in patients with COPD. *Diagnostics*, *9*(1), 1–13. https://doi.org/10.3390/diagnostics9010033
(VHB-JQ3: – | SJR: Q2 | Impact factor (2021): 3.992)
- Krupitzer, C., Drechsel, G., Mateja, D., Pollkläsener, A., Schrage, F., Sturm, T., Tomasovic, A., & Becker, C. (2018). Using spreadsheet-defined rules for reasoning in self-adaptive systems. In *Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)* (pp. 289–294). Athens, Greece. https://doi.org/10.1109/PERCOMW.2018.8480283

The rest of this dissertation is organized as follows. The overarching research context is discussed in Chapter 2. Next, the four papers that are included in this dissertation are presented in Chapters 3–6.[4] Finally, Chapter 7 concludes the dissertation with a discussion of the studies' overarching contributions and future research endeavors.

[4] The papers have been slightly adapted from their original versions to have a consistent layout throughout this dissertation. They are written from the first-person plural perspective (i.e., "we"), since several co-authors contributed to each publication.

2 Research Context and Positioning of this Dissertation

This chapter introduces the overarching concepts and theories that are used in the four included papers. These include AI and ML, coordination and organizational performance, and organizational theories that are central to the four studies (i.e., problem solving, task–technology fit, and organizational learning).[5]

2.1 Artificial Intelligence and Machine Learning

This section introduces the overarching concept of AI and discusses related themes. It then defines ML, a modern AI approach, and discusses its realization in organizations.

2.1.1 Artificial Intelligence

The term *artificial intelligence* is difficult to define. Since the birth of AI as an academic discipline in 1956 during the famous and influential *Dartmouth Summer Research Project on Artificial Intelligence* workshop (McCarthy et al., 1955, 2006), numerous definitions of AI have been (and are still being) proposed (e.g., Berente et al., 2021; Russell & Norvig, 2021; Wang, 2019). The continuing discussion of how to define AI appears to mainly result from an equally absent consensus on the definition of *intelligence*, whose definitional challenges are transmitted to AI, which aims to emulate intelligence (there are excellent overviews by, e.g., Legg & Hutter, 2007). This leaves the term *AI* without a clear definition, despite numerous redefinitions (e.g., Berente et al., 2021; Wang, 2019), proposed intelligence tests (e.g., Searle, 1980; Turing, 1950), and reflections (e.g., Benbya et al., 2021; Russell & Norvig, 2021).

Yet what does exist today is a substantial consensus on major concepts of artificial entities that are widely perceived as in some way behaving intelligently (e.g., Russell & Norvig, 2021; Schuetz & Venkatesh, 2020). One central concept is the notion of the *rational agent* proposed by Russel and Norvig (2021), which has been widely accepted by scholars across disciplines (e.g., Berente et al., 2021; Nilsson, 1998; Russell & Norvig, 2021; Schuetz & Venkatesh, 2020). According to this concept, an intelligent agent can be defined as "anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators" (Russell & Norvig, 2021, p. 54).[6] To this end, intelligent behavior is implemented as an agent function that aims to select the seemingly best action based on current context information (Russell & Norvig, 2021). For instance, a self-driving car constitutes a rational agent that is led by its agent function to make the most reasonable decision about its actions (e.g., accelerating, braking, steering)

[5] This section mainly presents a synthesis of the theoretical backgrounds provided in papers A to D. While major aspects are discussed here, please refer to the respective papers in Chapters 3–6 for further details.

[6] Note that "environment" and "actions" are loosely defined in this definition. For instance, a solution for image recognition also fulfills the definition of an intelligent agent. Here, rather counterintuitively, the environment is represented by a set of given images, and actions are a set of selectable labels to describe a given image (Russell & Norvig, 2021).

based on data about its driving environment (e.g., road conditions, the position of pedestrians, the car's velocity; Russell & Norvig, 2021). There exist various approaches to realizing agent functions (e.g., through manually defined rules or statistics; Russell & Norvig, 2021). Today, ML is a prominent and promising approach to realizing agent functions, as it enables ISs to learn agent functions on their own (e.g., Brynjolfsson & Mitchell, 2017; Russell & Norvig, 2021). As ML is an approach that has enabled recent AI advances (e.g., He et al., 2015; Silver et al., 2017; Vinyals et al., 2019) and spurred the current advent of AI in organizations (e.g., Benbya et al., 2021; Berente et al., 2021; Brynjolfsson & Mitchell, 2017), I focus on AI as an intelligent agent that relies exclusively on ML (see Section 2.1.2 for further details about ML) in this dissertation.

AI research distinguishes two major scopes of AI, namely, *narrow AI* and *artificial general intelligence* (AGI). *Narrow AI* refers to AI that can only perform a single task in a specific domain (Benbya et al., 2021; Brynjolfsson & Mitchell, 2017; Jordan & Mitchell, 2015). For instance, a single narrow AI cannot cover the three tasks of classifying images, recommending a song, and playing chess; rather, this requires building an individual AI for each of the three tasks (e.g., He et al., 2015; Liebman et al., 2019; Silver et al., 2017). While some narrow AIs are able to reach superhuman performance in their specialized tasks, such AI cannot be applied beyond the specific task and domain for which it was developed (de Bruyn et al., 2020). Narrow AI also encompasses the concept of *weak AI* (in contrast to *strong AI*)—that is, AI that can only simulate intelligent thinking but cannot perform the actual thinking process that occurs in brains (Searle, 1980). In contrast to narrow AI, *AGI* is a general-purpose AI (Benbya et al., 2021; Brynjolfsson & Mitchell, 2017; Raisch & Krakowski, 2021). AGI aims for a more generalist focus, that is, being able to perform a wide variety of complex tasks in various different domains. For example, a single AGI may clean rooms, design airplanes, and compose music—just as humans can. AGI is often assumed to also entail the concept of strong AI, that is, being able to perform the actual thinking process that occurs in brains instead of only simulating it (Searle, 1980). While AGI is still limited to the utopian and dystopian futures portrayed by science-fiction authors, narrow AI has already entered today's reality, changing our societies and economies (Brynjolfsson & Mitchell, 2017; Jordan & Mitchell, 2015; Russell & Norvig, 2021). Since it is unclear when and whether humanity will ever actually create AGI, I focus my analyses and contributions in this dissertation on narrow AI and thereby also adopt the focus of current research on organizational AI (e.g., Benbya et al., 2021; Berente et al., 2021; Brynjolfsson & Mitchell, 2017; de Bruyn et al., 2020; Raisch & Krakowski, 2021).

2.1.2 Machine Learning

ML algorithms enable ISs to derive patterns from data, which are then used to create ML models (Brynjolfsson & Mitchell, 2017; Jordan & Mitchell, 2015; Mitchell, 1997; Russell & Norvig, 2021). As ML models can act as solutions to real-world problems, ML represents a new programming paradigm: In traditional ISs without ML, human experts must solve given problems and then translate their solutions into code to craft ISs for certain tasks (e.g., humans defining rules to design a vacuum robots' fixed routines; Russell & Norvig, 2021). With ML, it is not human experts but the ML systems themselves that define solutions based on the patterns they derive from data (e.g., ML systems learning how to detect credit card fraud from business transactions; Ala'raj & Abbod, 2016; Kruppa et al., 2013). By doing so, ML can render manual programming obsolete for certain tasks (Samuel, 1959).

Three major types of ML can be distinguished (Bishop, 2006; Mitchell, 1997; Russell & Norvig, 2021). First, *supervised learning* aims to learn a model from a collection of input–output pairs (i.e., labeled data; Russell & Norvig, 2021) that can assign output to new input data (Bishop, 2006; Russell & Norvig, 2021). Examples of supervised learning tasks include classification and regression problems (i.e., assigning nominal or numerical outcomes to input data; Bishop, 2006; Mitchell, 1997) that underlie applications such as image classification and revenue prediction (e.g., Brynjolfsson & Mitchell, 2017). Second, *unsupervised learning* aims to learn a model that captures commonalities in known data inputs (i.e., unlabeled data; Russell & Norvig, 2021), allowing a response to the absence or presence of the learned commonalities in new input data (Bishop, 2006; Russell & Norvig, 2021). As an example, clustering represents a classic unsupervised learning problem (Bishop, 2006) that enables, for instance, autonomously grouping documents with similar topics (e.g., Jain et al., 1999). Third, in *reinforcement learning,* an agent aims to interact with its environment to learn a policy for choosing future actions that maximize the rewards received (Bishop, 2006; Russell & Norvig, 2021). Application examples include autonomous driving (e.g., Michels et al., 2005), trading (e.g., Dempster & Leemans, 2006), and playing (video) games (e.g., Silver et al., 2017; Togelius et al., 2009). While those three types of ML are currently prevalent in theory and practice, hybrid approaches and new types of ML have also been discussed (e.g., Berthelot et al., 2019; Lan et al., 2020). In this dissertation, I aim to abstract the derived contributions beyond specific ML types, with the hope of informing research on all types of ML.

ML systems are developed iteratively (e.g., Kurgan & Musilek, 2006; Wirth & Hipp, 2000). Humans choose a problem, select and prepare problem-related data, select and parametrize ML algorithms, evaluate the implemented alternatives, and iterate between these tasks to eventually craft the best-performing ML system possible. By doing so, humans define the conditions under which an ML system learns to develop its own understanding of a problem solution (e.g., Amershi et al., 2019; Sturm, Gerlach, et al., 2021). Humans continually repeat these actions to reconfigure ML systems as the problem or available knowledge about it changes over time. The behavior of the resulting ML systems therefore strongly depends on the understanding of problems and anticipated solutions of humans involved in the development process. The more those who are involved know about the problem, the better they can set up the ML, and the higher the ML system quality will be. In contrast, a lack of expertise on the part of the involved humans can introduce incorrect or biased ideas about a problem into an ML system (Choudhury et al., 2021; Diakopoulos, 2016; Schuetz & Venkatesh, 2020). Management of the human expertise that is used to craft and reconfigure ML systems should thus be a major concern for organizations when using ML systems. In this dissertation, I therefore explore the effects of the initial creation (see papers A, B, and D) and the continual reconfiguration (see papers A, C, and D) of ML systems with special attention paid to the human expertise that is involved.

2.1.3 Human–AI Collaboration

Traditionally, ISs were regarded exclusively as a tool to support humans in completing their tasks (e.g., Alavi & Leidner, 2001; Goodhue & Thompson, 1995; Kane & Alavi, 2007). For instance, email and knowledge repositories are used to support human communication and store human knowledge rather than communicating and creating their own knowledge (e.g., Kane & Alavi, 2007). With ML systems' capability to learn their own problem solutions, ML systems challenge this assumption as they have greater autonomy in executing tasks, thus shifting ISs into more autonomous roles in organizations (e.g., Berente et al., 2021; Schuetz & Venkatesh, 2020). This has stimulated a broad discussion on whether and when ML systems should be used to automate or augment human task execution (e.g., Brynjolfsson & Mcafee, 2016; Daugherty & Wilson, 2018; Davenport & Kirby, 2016). In particular, the shrinking role of humans and the threat of their replacement have become widely discussed topics of concern (e.g., Brynjolfsson & Mitchell, 2017; Faraj et al., 2018; Tschang & Almirall, 2021). Yet the automation–augmentation narrative is not without flaws. In their influential paper, Raisch and Krakowski (2021) show that the distinction between automation and augmentation involves a fundamental paradox: When enlarging the perspective along space and/or time, an ML system that appears to automate a task may also augment a task at the same time (or vice versa) without changing anything about the use of the ML system, which simultaneously places the ML system into automating and augmenting roles (e.g., one subtask may be fully automated by an ML system which may then be combined with other human-performed subtasks to augment an overall task). Thus, after years of discussion and more available empirical evidence, the narrative has once again recently shifted, and humans and ML systems are now viewed in major discussions on AI as counterparts that collaborate with each other in organizations (e.g., Fügener, Grahl, Gupta, & Ketter, 2021; Kane et al., 2021; Raisch & Krakowski, 2021; Ransbotham et al., 2020; Seidel et al., 2019; Sturm, Gerlach, et al., 2021; Teodorescu et al., 2021; Tschang & Almirall, 2021). In this dissertation, I therefore adopt the view that humans and ML systems can act as collaborators that aim to jointly achieve organizational objectives. Abstracting from the automation–augmentation distinction allows me to focus my analyses on the core change behind the use of ML systems, which is that ML systems do not simply support humans, but rather, both humans and ML systems are now able to create and share their own contributions within the organization. In doing so, I join recent research on organizational AI that aims to advance our understanding of the role of humans and ML systems (e.g., Ågerfalk, 2020; Fügener, Grahl, Gupta, & Keter, 2021; Grønsund & Aanestad, 2020; Jussupow et al., 2021; Kane et al., 2021; Raisch & Krakowski, 2021; Ransbotham et al., 2020).

2.2 Coordination and Organizational Performance

An organization is not a single entity with its own objectives, desires, and abilities. Instead, an organization constitutes a complex system of multiple interacting agents that aim to jointly define and work towards collective goals (e.g., March, 1991; Nickerson & Zenger, 2004). If done correctly, the effective collaboration of the agents is worthwhile as it allows more than just the sum of the individuals' successes to be achieved through exploiting achievable synergies through their interactions (e.g., Lyles & Fiol, 1985; March, 2010). This requires the agents to agree on the same goals and support one another, as well as having agents work on the tasks that fit their abilities best (e.g., Argote et al., 2021; Argote & Miron-Spektor, 2011; Grant, 1996; Levitt & March, 1988; March, 1991). When agents are well organized, virtuous dynamics can emerge that help foster the agents' abilities to jointly exceed the organization's objectives (e.g., Argote et al., 2021; Argote & Miron-Spektor, 2011; March, 1991; K. D. Miller et al., 2006). However, if agents are ill-organized, they may pursue contradictory goals and hinder one another's actions, potentially creating vicious dynamics and detrimental chaos (e.g., Argote et al., 2021; Argote & Miron-Spektor, 2011; Levinthal & March, 1993; March, 1991, 2006). The means to organize agents' dynamics is effective coordination, that is, "the act of making all the people involved in a plan or activity work together in an organized way" (University of Cambridge, 2022). Unfortunately, it is not easy to achieve effective coordination (e.g., Levinthal & Rerup, 2021; March, 1991; Weick et al., 2005). Agents' peculiarities and manifold interactions render foreseeing and controlling emerging dynamics a very complex endeavor (e.g., Benbya et al., 2020; Grant, 1996; Levinthal & March, 1993). For this reason, although research has analyzed coordination for decades, research continues to study its manifold consequences and mechanisms (e.g., Argote et al., 2021; Argote & Miron-Spektor, 2011; Huber, 1991; Lyles & Fiol, 1985).

Organizational performance reflects the efficiency, effectiveness, and/or quality of task performance in organizations, revealing organizations' productivity (Goodhue & Thompson, 1995). Performance is vital to organizations, as poorly performing organizations cannot survive for long in a competitive environment (e.g., Levinthal & March, 1993; March, 1991). Coordination is strongly connected to organizational performance (e.g., Grant, 1996; March, 1991). The better an organization's agents are coordinated, the better the agents' actions are integrated and complemented, which increases the organization's performance (e.g., Grant, 1996; March, 1991). Along the three levels of analysis pursued in this dissertation, three types of performance can be distinguished: individual-, group-, and organization-level performance. The connection between the three types of performance is nontrivial. Effective coordination that increases individual-level performance may at the same time be detrimental to group- or organization-level performance (and vice versa; e.g., Fang et al., 2010; March, 2006; Schilling & Fang, 2014). For instance, while sharing unique expertise with teammates may reduce an individual's relative performance, it may increase the group performance (e.g., Fang et al., 2010; March, 2006). This prevents a simple translation of findings between levels, making separate analyses of the three levels necessary.

2.3 Theoretical Foundations

To analyze the coordination of humans and ML systems along the three levels of analysis, I apply different theoretical perspectives in this dissertation. For the individual level, I first rely on a problem-solving perspective to unpack the notion of ML systems as a new form of organizational problem solvers. Here, problem-solving research helps inform the characterization of problems and the process of identifying and evaluating problems that are suited to be solved by ML systems. Next, to emphasize interactions, I turn to research on TTF. TTF theory offers a powerful theoretical framework for capturing the interplay between an individual and a technology with regard to individual performance. The use of this framework helps uncover the conditions for effective collaboration between a human and an ML system on a shared task. For analyzing group- and organization-level performance, I rely on an organizational learning perspective. Learning not only captures ML systems' unique capabilities but is also the essential process that connects humans' and ML systems' behavior, and organizational learning research provides a fruitful theoretical basis to help capture and explain the dynamics that arise from interactions between multiple agents—no matter whether the agents are humans or ML systems. In what follows, I explain the three applied theoretical perspectives in more detail.

2.3.1 Problem Solving

Problem solving is a crucial key activity in organizations (e.g., Björk, 2012; Lang et al., 1978; Lyles & Mitroff, 1980), as it stimulates organizational change through the generation of new, suitable ideas (i.e., problems, solutions, and solution implementations; e.g., Basadur et al., 1982; E. Mumford, 1998; Reiter-Palmon & Robinson, 2009; van den Ende et al., 2015). A widely accepted model that synthesizes activities involved in the problem-solving process (e.g., Lang et al., 1978; E. Mumford, 1998; Reiter-Palmon & Robinson, 2009; Smith, 1989) is the one by

Basadur et al. (1982) that is based on Leavitt's (1975) three-phase model: First, the phase of *problem finding* covers the recognition, identification, and construction of problems to make a set of potentially solvable problems available (Chi et al., 1981). Second, based on selected problems, the *problem-solving* phase focuses on searching suitable solutions (M. Mumford et al., 1994). Finally, the phase of *solution implementation* aims to craft implementations of selected solutions, integrating them into organizational processes (M. D. Mumford et al., 1997). Each of the three main phases further involves two subphases following Osborn's (1953) divergence-convergence dualism (Basadur et al., 1982; Osborn, 1953): First, *ideation* aims to uncover and collect ideas on potential problems, solutions, or implementations respectively. The subsequent *evaluation* screens identified ideas to distinguish good ideas from bad ones. If done right, the problem-solving process allows organizations to continuously improve over time, ensuring long-term success through effective solutions to substantial problems.

The problem-solving process offers a valuable perspective for analyzing the ML systems in organizations because ML systems, with their ability to derive their own solutions from data (Mitchell, 1997; Russell & Norvig, 2021), can be viewed as a new form of organizational problem solver. Since human problem finding remains relevant in ML contexts, as humans must still define the problems for ML systems (e.g., Amershi et al., 2019; Diakopoulos, 2016; Sturm, Gerlach, et al., 2021; see also Section 2.1.2), an exploration of the problem-finding phase can particularly help organizations reflect on relevant aspects and conditions when they use ML systems as problem solvers and gain insights regarding how to enable an effective delegation of problems to ML systems (e.g., Brynjolfsson & Mitchell, 2017; Fügener, Grahl, Gupta, & Keter, 2021). Paper A therefore adopts a problem-finding perspective to synthesize experts' notions of the nature of problems suited to be solved by ML systems and how to identify such problems.

2.3.2 Task–Technology Fit

Originally referred to as the "technology-to-performance chain," the now seminal TTF theory was proposed by Goodhue and Thompson (1995) to model the link between IT use and individual performance. In this theory, Goodhue and Thompson (1995) conceptualized a simple but powerful observation: To improve individual performance, the IT that is used must be a good fit for the tasks it supports. To nuance this observation, Goodhue and Thompson (1995) built the TTF theory from five main constructs:

- the *characteristics of tasks* that individuals perform to turn inputs into outputs;
- the *characteristics of technologies* that support individuals in performing the tasks;
- the *task-technology fit*, or how well a technology supports individuals' tasks;
- the *utilization* as the individual usage behavior of IT that is used to perform tasks; and
- the *performance impacts* on individuals' tasks as some combination of improved efficiency, effectiveness, and/or quality.

TTF theory indicates that (1) the *task* and *technology characteristics* jointly affect the *task-technology fit*, (2) increasing the *task-technology fit* tends to increase both the *utilization* of IT and individual *performance impacts*, and (3) increasing the *utilization* of IT can also increase the *performance impacts*. The *task-technology fit* can thus have both a positive direct and a positive indirect effect (through the mediating *utilization* construct) on *performance impacts*.

TTF theory has been used to analyze performance impacts in numerous technology contexts, such as group support systems (e.g., Dennis et al., 2001; Fuller & Dennis, 2009; Maruping & Agarwal, 2004; Zigurs & Buckland, 1998), mobile IS (e.g., Gebauer et al., 2010; Gebauer & Ginsburg, 2009; Junglas et al., 2008; Lee et al., 2007), expert systems (e.g., Wongpinunwatana et al., 2000), data analytics (e.g., Ghasemaghaei et al., 2017; Karimi et al., 2004), and decision support systems (e.g., Parkes, 2013). So far, TTF theory has not been translated to the context of ML systems. As TTF theory provides a powerful theoretical structure to analyze how the link between an individual and IT affects individual performance, paper B contextualizes TTF theory to capture ML systems' impact on individual performance.

2.3.3 Organizational Learning

In their seminal paper on organizational learning, Levitt and March (1988) view organizations as "learning by encoding inferences from history into routines that guide behavior" (p. 319). Following this notion, organizations continuously learn from past actions to gradually improve their future actions over time (e.g., Argote et al., 2021; Argote & Miron-Spektor, 2011; Levitt & March, 1988). As organizations must be able to effectively act and adapt to survive in their competitive environments, organizational learning is vital to organizations as it defines such adaptive organizational behavior (e.g., Argote & Miron-Spektor, 2011; Levitt & March, 1988; March, 1991, 2010). Yet organizations cannot learn by themselves, but must rely on the learning of their members (e.g., Huber, 1991; Levitt & March, 1988; March, 1991). Over time, organizational members accumulate experience through their individual actions, formulate beliefs based on their experience of chosen actions and associated outcomes, and share their developed knowledge with other individuals (e.g., Argote & Miron-Spektor, 2011). In this way, organizational learning is more than just the sum of individual learning, because the transfer of knowledge itself creates value (e.g., by stimulating further learning endeavors and combining knowledge; Argote & Miron-Spektor, 2011; Lyles & Fiol, 1985; Nonaka, 1994). Organizations utilize their members' learning by storing these in routines and knowledge repositories to apply and share the developed knowledge (March, 1991). Therefore, from an organizational learning perspective, an organization represents a complex system of interacting individuals who learn to collectively make sense of the organization's environment (Levitt & March, 1988). To achieve great learning performance, organizations must be able to effectively coordinate the individual and mutual learning of their members to avoid the emergence of detrimental chaos within the learning system (e.g., Grant, 1996; Lavie et al., 2010; March, 1991). Decades of research have emphasized that ineffective coordination of organizational learning can therefore have wide-ranging consequences for organizational performance and can even threaten organizations' long-term survival (e.g., Argote & Miron-Spektor, 2011; Gupta et al., 2006).

One of the most essential concepts in coordinating organizational learning is the distinction between exploitative and explorative learning (e.g., Gupta et al., 2006; Lavie et al., 2010; March, 1991; Raisch et al., 2009). *Exploitation* focuses on the incremental search for and refinement of ideas, aiming to remain in the reliable near neighborhood of extant knowledge. In contrast, *exploration* aims to shift away from extant knowledge by searching for unorthodox ideas, attempting to "look outside the box" (e.g., Gupta et al., 2006; Lavie et al., 2010; March, 1991). March (1991) showed that the secret for a high level of long-term effectiveness lies in *balancing exploitation and exploration*—a notion that has spurred thousands of studies confirming and nuancing March's essential observation (e.g., Argote et al., 2021; Argote & Miron-Spektor, 2011; Gupta et al., 2006; Lavie et al., 2010; March, 1991; Raisch et al., 2009). If organizations strongly exploit ideas but neglect exploring, they risk becoming trapped in stagnation and failing to continue to adapt. If organizations explore extensively but do not exploit, they will fail to develop specific competences that are required to survive in competitive environments (e.g., Gupta et al., 2006; March, 1991; Raisch et al., 2009).

Finding a good balance between the two processes of exploitation and exploration has proved to be a serious challenge for research and practice. Numerous flaws of organizational learning have made it difficult for organizations and scholars to put forth a universal solution for achieving a well-balanced state of great organizational performance (e.g., Argote et al., 2021; Argote & Miron-Spektor, 2011; Gupta et al., 2006; Levinthal & March, 1993; March, 2006, 2010; Raisch et al., 2009). One of the most discussed flaws is *learning myopia*, that is, the tendency to favor exploitation over exploration (Levinthal & March, 1993). Due to the uncertain benefits of experimenting with new and ambiguous ideas, learners prefer to utilize ideas that have already proved to be reliable in the past (e.g., Levinthal & March, 1993; March, 2006). There exist various factors that tend to either promote (e.g., rewarding successes and penalizing failures; March, 2010) or alleviate (e.g., increasing team diversity; March, 1991) learning myopia, adding to the complexity that is involved in achieving the crucial exploration–exploitation balance (e.g., Levinthal & March, 1993; March, 1991, 2010).

Organizational learning has been analyzed in numerous contexts (excellent overviews exist; see, e.g., Argote et al., 2021; Argote & Miron-Spektor, 2011; Huber, 1991). Yet research on organizational learning in the IS context is still in its infancy (Argote et al., 2021; Argote & Miron-Spektor, 2011). To date, there exist only a handful of studies on organizational learning in the IS context (e.g., Balasubramanian et al., 2022; Dodgson et al., 2013; Kane & Alavi, 2007), even though the potential of IS to support learning and the accumulation of knowledge has been widely acknowledged (Alavi & Leidner, 2001; Argote et al., 2021; Argote & Miron-Spektor, 2011). The role of ISs has always been assumed to be exclusively passive, with ISs only being able to support human learning (e.g., Argote & Miron-Spektor, 2011). With the rise of ML systems and given their unique ability to learn, ML systems can contribute their own learning to organizational learning processes, increasingly shifting ML-based ISs towards a more active role (e.g., Argote et al., 2021; Lindebaum et al., 2020; Mateja & Heinzl, 2021; Ransbotham et al., 2020). Despite the rising role of ML systems, research on ML systems in organizational learning remains scarce (i.e., Afiouni-Monla, 2019; Balasubramanian et al., 2022; Lyytinen et al., 2021; Ransbotham et al., 2020; Seidel et al., 2019). Insights about organizational learning may help to further unpack ML systems' unique learning capabilities and may offer an informative link to human behavior (e.g., Lyytinen et al., 2021; Ransbotham et al., 2020; Seidel et al., 2019). In addition, organizational learning provides powerful theory and research instruments that may help to unfold and explain the complex dynamics between humans and ML systems (e.g., Argote et al., 2021; Lyytinen et al., 2021; Sturm, Gerlach, et al., 2021). For instance, organizational learning research has contributed a rich set of conceptualized learning patterns (e.g., learning myopia or superstitious learning; Levinthal & March, 1993; March, 2010) and simulation models (e.g., Gavetti & Levinthal, 2000; Levinthal, 1997; March, 1991). This allows abstracting from complex dynamics and focusing analyses on relevant variables, helping to better explain the evolution of dynamics within the complex context of learning (e.g., Balasubramanian et al., 2022; Sturm, Gerlach, et al., 2021). Papers C and D therefore use the organizational learning perspective to help observe and unpack humans' and ML systems' behaviors and to coordinate their intertwined learning dynamics.

3 Paper A: Artificial Intelligence and Individual-level Performance (Focus: Problem Solving)

Title

To Use or Not to Use Artificial Intelligence? A Framework for the Ideation and Evaluation of Problems to Be Solved with Artificial Intelligence

Authors

Sturm, Timo; Fecho, Mariska; and Buxmann, Peter

Publication Outlet

Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS 2021)

Abstract

The recent advent of artificial intelligence (AI) solutions that surpass humans' problem-solving capabilities has uncovered AIs' great potential to act as new type of problem solvers. Despite decades of analysis, research on organizational problem solving has commonly assumed that the problem solver is essentially human. Yet, it remains unclear how existing knowledge on human problem solving translates to a context with problem-solving machines. To take a first step to better understand this novel context, we conducted a qualitative study with 24 experts to explore the process of problem finding that forms the essential first step in problem-solving activities and aims at uncovering reasonable problems to be solved. With our study, we synthesize emerged procedural artifacts and key factors to propose a framework for problem finding in AI solver contexts. Our findings enable future research on human-machine problem solving and offer practitioners helpful guidance on identifying and managing reasonable AI initiatives.

Keywords

Artificial intelligence, ideation, innovation, machine learning, problem solving

3.1 Introduction

In recent years, advances in artificial intelligence (AI) allowed machines to master problems previously dominated by humans: AIs defeated world's best human GO player (Silver et al., 2017), recognized images better than the average human (He et al., 2015), and beat some of the greatest human StarCraft II players (Vinyals et al., 2019). Due to success stories like these, more and more organizations aim to explore how to use AIs' disruptive potential to improve their organizational performance (e.g., Bean, 2019; Forbes Insights, 2018; Schmelzer, 2019).

The technology that underlies such modern AI information systems (IS) is machine learning (ML) (Brynjolfsson & Mitchell, 2017; Jordan & Mitchell, 2015). Such ML-based AIs use ML algorithms to derive patterns from data and apply these patterns to new data to perform actions (Brynjolfsson & Mitchell, 2017; Mitchell, 1997; Russell & Norvig, 2021). ML thus constitutes a new programming paradigm: With ML, algorithms derive solutions from data, instead of having humans manually solving problems and translating their solutions into code anymore (Samuel, 1959). The resulting handover of problem-solving activities to data-driven algorithms therefore requires us to reassess the role of IS in organizations and our knowledge on how to manage IS successfully (e.g., Rai et al., 2019; Rzepka & Berger, 2018).

One process that essentially drives and ensures an organization's progress and is thus crucial for its long-term survival is the act of problem solving (Kolb, 1976; Lyles & Mitroff, 1980); that is, the act of finding, solving, and implementing solutions for problems (Basadur et al., 1982; H. J. Leavitt, 1975). For decades, scholars from various disciplines have analyzed problem solving from different perspectives (e.g., Delbecq & van de Ven, 1971; Lang et al., 1978). Yet, such organizational studies have commonly assumed that the solver of organizations' problems is only *human*. With ML-based AI essentially representing a technology for machine-driven problem solving that organizations increasingly adopt (Brynjolfsson & Mitchell, 2017; Mitchell, 1997; Russell & Norvig, 2021), this core assumption must be fundamentally questioned.

To take a first step to better understand how organizations can manage problem solving in the AI age, we explore how problem finding, which precedes the core problem solving activity and aims to identify relevant problems, translates to contexts where AIs act as problem solvers. To achieve this, we conducted a qualitative study with 24 experts that frequently conduct AI initiatives. We thus aim to answer:

(1) How can organizations find problems that are likely suited to be solved by ML-based AIs, and (2) which central factors likely determine ML-based AIs' suitability for solving a problem?

3.2 Theoretical Background

Below, we first present key concepts and related work of problem solving and AI. We then integrate both research streams to form our study's objective.

3.2.1 The Process of Problem Solving

Problem solving, the act of uncovering problems and searching for effective solutions (Hippel & Krogh, 2016; Simon & Newell, 1971; Smith, 1989; Taylor, 1975), is considered a key activity in organizations (Lang et al., 1978; Lyles & Mitroff, 1980). This process involves generating new, suitable ideas (i.e., both problems and solutions) at its core (Basadur et al., 1982; E. Mumford, 1998; Reiter-Palmon & Robinson, 2009), playing an essential role for organizations (Björk, 2012) as ideas stimulate organizational change (van den Ende et al., 2015).

One model that represents a solid synthesis of the widespread consensus of general stages of problem-solving processes (e.g., Lang et al., 1978; E. Mumford, 1998; Reiter-Palmon & Robinson, 2009; Smith, 1989) is the one originally proposed by Basadur et al. (1982) based on Leavitt's (1975) suggested tripartite model. Figure 1 illustrates this process that we describe below.

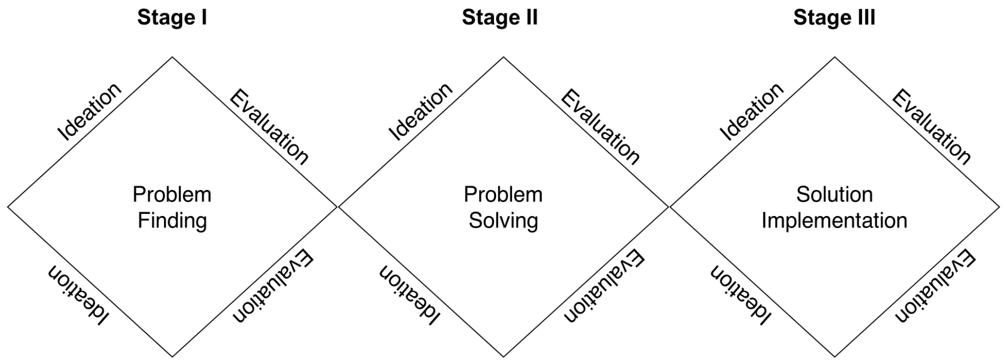

Figure 1: Problem-Solving Process (Basadur et al., 1982)

The problem-solving process comprises three steps: First, the initial *problem finding* aims to recognize, identify, and construct problems. Then, the *problem solving* includes the search for suitable solutions by exploring potentially fitting solutions for given problems (M. Mumford et al., 1994). Finally, the *solution implementation* pursues to integrate selected solutions into organizational processes (M. D. Mumford et al., 1997). Following Osborn's (1953) widely adopted divergence-convergence dualism (i.e., uncover choices and screen choices), each of the three above-mentioned stages also comprises a two-step subprocess (Basadur et al., 1982; Osborn, 1953): First, *ideation* aims to explore ideas. Depending on the three stages in Figure 1, this respectively refers to uncovering and constructing potential problems, solutions, and solution implementations. The subsequent *evaluation* then assesses the respective ideas that yielded from the ideation to distinguish good ideas from bad ones.

This process helps us to better understand the essential role of *problem finding*. Within this step, humans identify and construct problems on the basis of domain knowledge and prior experiences (Chi et al., 1981) to uncover problems together with related goals, possible problem-solving approaches, and restrictions (Reiter-Palmon & Robinson, 2009). Since ill-defined problems can contain characteristics that lead to unexpected or unsatisfactory results (e.g., selecting unsuitable approaches or missing relevant aspects), problem finding essentially affects the success of problem-solving activities (Hippel & Krogh, 2016; Simon, 1973).

For decades, scholars have examined how human problem solving, and in particular problem finding, translates to different contexts, such as the individual or group-level (e.g., Delbecq & van de Ven, 1971; Lang et al., 1978). In recent years, research has also started to examine problem finding for creating solutions with digital technologies. As a result, several frameworks were proposed that focus on digital technologies in general (e.g., Benta et al., 2017; Bremser, 2018; Kayser et al., 2018). For instance, most closely related to our study, Vanauer et al. (2015) propose an ideation framework for Big Data solutions. They found multiple procedural artifacts that comprise two ideation alternatives and several suitability assessments. Although ML-based AI also represents a digital technology, the existing frameworks neglect particularities that result from the unique problem-solving capabilities of ML-based AI.

3.2.2 Artificial Intelligence & Machine Learning

A widely used conceptualization of AI is the intelligent agent; that is, anything that can perceive context information and autonomously act upon that through actuators (Russell & Norvig, 2021). The technique that organizations increasingly use to implement such agents' behavior is ML (Brynjolfsson & Mitchell, 2017): Intelligent agents based on ML—by us referred to as *ML-based AI*—are based on algorithms that can identify patterns in data and use these patterns to act on new data (Mitchell, 1997). Without ML, humans solve problems manually and codify their solutions into traditional non-ML IS. In contrast, ML-based AIs derive their own solutions for defined problems exclusively from data, rendering manual programming unnecessary (Brynjolfsson & Mitchell, 2017; Samuel, 1959). While artificial general intelligence remains beyond reach, organizations increasingly use ML-based AI to solve narrow problems (Brynjolfsson & Mitchell, 2017), sometimes achieving solutions that even surpass human problem-solving capabilities (e.g., He et al., 2015; Silver et al., 2017; Vinyals et al., 2019). Especially in contexts where tasks comprise a limited execution clarity, ML-based AI offers a great potential to explore available alternatives and evaluate their properties more extensively and precisely than their human counterparts that are more limited in their information processing capabilities (Brynjolfsson & Mitchell, 2017).

To exploit this potential, organizations must understand how they can use ML-based AI to solve problems within their organizational contexts. Yet, existent research has only partially unveiled how organizations can use ML-based AI for their problem-solving activities. Thus far, scholars have intensively focused on understanding AI-driven problem solving to develop solutions for given problems: Professionals select and prepare data and also select and parametrize ML algorithms to frame a given problem and restrict potential solution designs. Next, they let the algorithms derive possible ML-based AI solutions and then evaluate the solutions. The professionals iteratively perform this process to eventually identify the best derivable AI solution (Brynjolfsson & Mitchell, 2017; Domingos, 2012; Mitchell, 1997). Existing research has proposed multiple frameworks to capture this process from different perspectives (e.g., Amershi et al., 2019; Domingos, 2012; Kurgan & Musilek, 2006).

Although this research generally expects problems to exist that must be solved, only a few studies stand out that explore how organizations can actually uncover problems to be solved with ML-based AI. The most applicable study is the one of Brynjolfsson and Mitchell (2017), in which they name basic criteria to identify suitable tasks for applying ML-based AI, but do not provide procedural guidance and only consider a single ML type (i.e., supervised ML). Besides, existing studies either regard this topic from a more strategic perspective to provide factors for ML-based AI adoption (e.g., Kruse et al., 2019; Pumplun et al., 2019; Sturm & Peters, 2020; Traumer et al., 2017), or focus on separate AI particularities, such as research on fair (e.g., Afrashteh et al., 2020; Martin, 2019; Rhue, 2019) and transparent AI (e.g., Diakopoulos, 2016; T. Miller, 2019; Peters et al., 2020).

3.2.3 The Need to Revisit Problem Finding

The problem-solving process by Basadur et al. (1982) offers a solid basis for exploring problem-solving activities in different contexts. Despite decades of research, humans have generally been considered the only actor that performs the second step within this process; that is, deriving appropriate solutions. As we are interested in understanding how the initial problem finding step translates to contexts where solutions are created by ML-based AI, we must reassess the fit of existent research with this novel context.

Moreover, research on the use of ML-based AI in organizational contexts has widely neglected the process of problem finding. Although AI research has conceptualized the development of ML-based AI solutions, the act of finding suitable problems has been widely overlooked so far. While some exceptions exist, corresponding studies do not provide procedural artifacts and mostly provide factors for more abstract or specific areas. Lastly, while problem-finding frameworks for digital technologies exist, they neglect to include relevant ML-based AI's particularities due to their divergent, more general technological focus.

So far, we miss insights to sufficiently explain how organizations can perform the initial, operative act of finding problems to be solved with ML-based AI. We therefore decided to conduct an explorative study to gather first evidence and propose a basic framework for problem finding within this novel context.

3.3 Qualitative Research Methodology

To explore central factors that influence the suitability of organizational problems for being suited to be solved with ML-based AI, we applied a qualitative research approach. In particular, we interviewed experts from the operational and management levels of different organizations that are highly involved in AI initiatives (Flick, 2004). We then followed the steps of a directed content analysis (Hsieh & Shannon, 2005) to contextualize problem finding for ML-based AI. According to Weber (1990), content analysis can be used to categorize and evaluate qualitative data.

Based on the proposed principles by Sarker et al. (2013), we formulated a semi-structed interview guideline that we used to lead the interviews. A high degree of coherence was ensured by discussing our definition of ML-based AI and selected use cases with each expert before every interview. We used semi-structured questions as they ensure that all relevant questions are posed, while allowing the experts to freely share own experiences and opinions (Myers & Newman, 2007). To examine various factors and procedural artifacts, our interview questions followed both an organizational and technological perspective to examine the essential ideation-evaluation process underlying problem finding (Basadur et al., 1982). Finally, our interview guide covered the following five sections: general information about the experts, ML-based AI particularities, organizational and technical requirements, identification and evaluation of AI usage scenarios, and potential benefits and risks related to the adoption of AI in organizational processes. The iterative approach during the interviews allowed a continuous adaptation of the initially defined questions. Thus, on the one hand, the focus of the investigation could be sharpened while, on the other hand, individual perceptions could be considered (Myers & Newman, 2007).

We selected interview partners, who have detailed experience in solving organizational problems with ML-based AI. We conducted 23 interviews with 24 experts from Europe and North America, including nine experts from user firms (i.e., that mainly purchase AI products) and 15 experts from provider firms to comply with data triangulation rules (Flick, 2004). One interview included two experts. The experts cover developers, data scientists, managers, pre-sales consultants, and technical consultants, who regularly deal with the design and integration of prototypical or productive systems in different organizational contexts. We noticed that we reached theoretical saturation (Flick, 2004) during the last five interviews as they yielded no further insights and thus stopped interviewing.

The interviews were conducted either by telephone or face-to-face between December 2018 and April 2019. On average, each interview lasted 56 minutes. They were recorded, transcribed, and analyzed using the NVivo 12 software. In line with Saldaña (2009), we performed an iterative multi-cycle coding process consisting of two coding cycles. The first cycle covered three types of coding: First, we used (1) attribute coding to extract essential information about participants and organizations. Then, we employed (2) hypothesis coding to determine and structure potential factors along insights of human problem finding. This step allowed us to stimulate code derivation by objectives and approaches of human problem finding that might be generalizable or ill-suited to the AI context. Finally, we applied (3) descriptive coding to identify new procedural artifacts and key factors that might extend the initial problem-finding process, allowing us to uncover AI-related particularities more independently from human problem finding. Since the first coding cycle resulted in a large number of factors (i.e., 11 procedural artifacts and 37 key factors), we used pattern coding in a second cycle to cluster similar constructs to form mutually exclusive and collectively exhaustive procedural artifacts and key factors.

This two-cycle coding process was performed individually and independently by two of the authors and two student assistants. All results were compared and integrated in discussions with all four parties until a consensus was reached: an initial framework emerged in which procedural artifacts were integrated into phases of problem finding and each procedural artifact was fortified with associated key factors. To achieve research rigor, the coding process and initial framework were validated in subsequent discussions between five IS researchers and three student assistants (Flick, 2004). Additional data sources regarding problem finding for digital technology contexts and the use of AI in organizations (i.e., Agrawal et al., 2017; Benta et al., 2017; Bremser, 2018; Brynjolfsson & Mitchell, 2017; Fedyk, 2016; Kayser et al., 2018; Satell, 2018; Traumer et al., 2017; Vanauer et al., 2015) were also considered to compare results with existing knowledge. Based on this data and investigator triangulation (Flick, 2004), the final framework was formed.

3.4 Results

With our study, we explored how problem finding translates to a context, in which the problem is aimed to be solved with ML-based AI. Our results show first evidence for fundamental procedural artifacts and related key factors. In our interviews, it got apparent that the fundamental phases of problem finding in a human solver context (i.e., ideation and evaluation) remain valid for an ML-based AI context. Moreover, we found that finding a problem for being solved with ML-based AI is determined by three central aspects; that is, a clear organizational purpose, available data, and technical particularities of ML-based AI. A model for problem finding in an ML-based AI context should therefore essentially follow an ideation and evaluation phase, while considering subphases driven by factors of the three central aspects.

Figure 2 illustrates the framework that emerged from our interviews. Its main structure follows a two-phase character: An ideation phase first aims to uncover potential problems. Within this phase, problems are explored aiming to fulfill both a clear organizational purpose and availability of required data. The subsequent two-step evaluation phase then aims to assess the suitability of envisioned problems. The phase starts with evaluating problem substance of uncovered problems; that is, to generally be suited for ML-based AI. If this essential suitability can be confirmed, evaluating problem particularities follows to clarify whether and to which extent special features of ML-based AI solutions fit given problems.

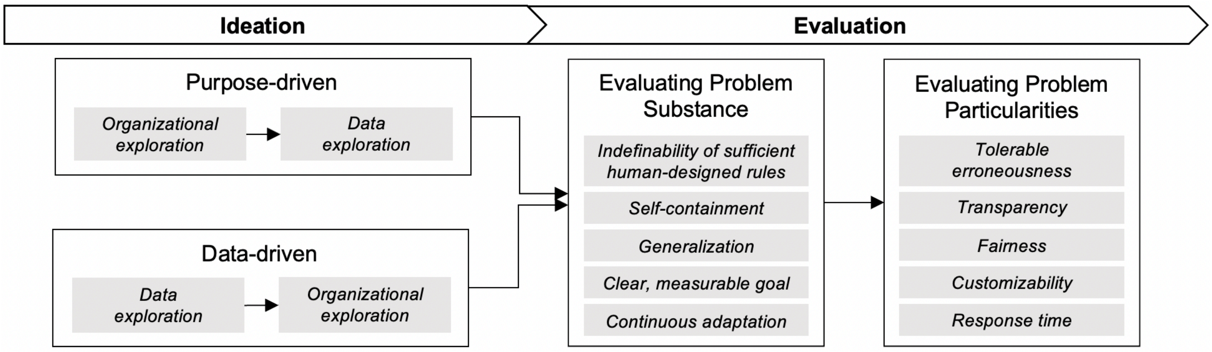

Figure 2: A framework for the ideation and evaluation of problems to be solved with ML-based AI

Note that this evaluation phase focuses on narrowing down problems based on their suitability for being solved with ML-based AI from a technical point of view, which our experts deemed central to problem evaluation in the AI context. As with human problem finding, construct and content-wise organizational evaluations likely extend our conceptualized evaluation phase to further assess problems' adequacy for being solved with respect to an organization's objectives. As this is out of our study's scope, we leave it to future research to analyze potential aspects for such problem evaluation foci while we abstract these hereinafter.

Below, we present the derived concepts. As each concept is grounded on some degree of consensus between the experts, we also indicate this degree as percentage of experts focusing each respective concept.

3.4.1 Ideation Phase

Within our interviews, it got apparent that a worthwhile ML-based AI solution unites a reasonable organizational purpose, available required data, and the fulfillment of ML-based AIs' technical particularities. Otherwise, even if the defined problem is basically suited to be solved with ML-based AI, an implemented solution may end up not being used if no clear organizational purpose is included or the implementation may fail if the required data turns out to be insufficient to create a functioning solution.

To increase the likelihood of finding suitable problems already in an early stage, the ideation phase should involve three essential elements: organizational exploration, data exploration, and an AI-specific problem substance. While the organizational and data exploration guide where the organization aims its ideation, the problem substance defines what the organization is searching for; that is, a set of factors that must be fully satisfied. We introduce them as part of a core assessment of the evaluation phase in section 3.4.2.1. Yet, these factors are also used to already guide the ideation. We further found that such ideation can be performed either in a purpose- or data-driven manner, which differ in the two explorations' order and focus. Below, we first introduce the two explorations and then integrate them into two ideation alternatives.

3.4.1.1 Organizational Exploration

Our interviews yielded that problem finding for ML-based AI should involve an exploration of organizational contexts to identify problems with a real and relevant organizational purpose. If such a purpose for an ML-based AI solution is missing, the experts emphasize that it becomes unlikely that the solution adds any value to the organization. While this may seem obvious, the majority of experts (63%) also highlight that organizations often fail to question the added value of having ML-based AI solutions for their problems.

To face this issue, the experts point out that organizations can actively search for organizational purposes by pursuing two alternative trajectories: First, organizations can focus on exploring how ML-based AI may **(1) replace existing solutions** (79%). In this case, organizations revisit established routines and offerings to explore whether ML-based AI may offer better ways to solve underlying problems. The experts agree that ML-based AIs can often offer valuable alternatives especially in domains in which designed solutions are bounded by humans' limited information processing capabilities or involved safety issues, or when their execution covers extensive manual efforts:

> *"'Do I need to increase profit by selling more or do I have a lot of production costs?' Such analysis needs to be done. Whether it's manufacturing, banking, or retail, I need to know what my business is and where I can improve given solutions. Then, we can focus on how ML can help." (i9)*

Second, organizations can use ML-based AI to **(2) explore new problem domains** (71%) to form entirely new offerings or routines. In this case, organizations can use ML-based AI as a driver to uncover problems that were previously out of the organization's scope. While this certainly includes to explore completely new processes and offerings, the experts stress that organizations can often benefit from AI through revisiting problems that were previously unsolvable due to manual, technical, or economic limitations:

> *"I think you get two broad categories of either: trying to solve a problem that rendered unsolvable so far or looking for new opportunities that you didn't know existed by analyzing your environment in an open-minded way." (i10)*

3.4.1.2 Data Exploration

An ML-based AI's possible solution space and thus framing of potential problems is basically determined by consumable data. The experts (88%) thus state that organizations can also guide their ideation by exploring usable data, using different foci: First, organizations can uncover potentially frameable problems along their data availability. The experts state that organizations usually start exploring data already available in electronic form. Then, they widen their focus to further sources like existing analog, public, purchasable, user, or newly recordable data (e.g., through adding novel sensors). The experts emphasize that organizations should pay special attention to high-volume data sources as this may imply a greater extent of potentially captured problem instances and thus a higher likeliness of uncovering representative data bases. Yet, the experts further stress that this allows organizations to especially uncover problem domains for which rather non-exhaustive, manually performed solutions exist due to humans' limited information processing capabilities that impede the analysis of extensive data volumes in a precise, comprehensive manner. In any case, the experts emphasize that organization must consider internal and external access restrictions (e.g., privacy issues or data ownership) to identify any access gaps as early as possible:

> *"'What data can we use?' Sounds trivial, but this is a huge problem. [...] Often, organizations do not even know the data they have, or it is distributed over so many systems that it would take*

forever to gather it. [...] What further data can we collect? What data can we additionally purchase?" (i4)

3.4.1.3 Purpose- & Data-driven Ideation

While organizational and data exploration should be covered eventually, our interviews indicate that both represent alternative starting points for the ideation phase. For both approaches, organizational and data exploration represent necessary investigations to ensure clarification and inclusion of an organizational purpose and required data. However, depending on the selected approach, the experts also agree that the two exploration types are used for a different purpose; that is, either for the initial identification of potential problems or the respective subsequent exploration of a fitting organizational purpose or usable data.

The experts (92%) state that, when following a **(1) purpose-driven ideation**, potential problems are initially derived from an organizational context. For each derived problem, potentially usable data is then explored to grasp the available technical foundation. The experts further stress that, if the required data is not available and cannot be made available in a reasonable manner, the identified potential problem should be dismissed or postponed to be solved with AI:

"The customer must have a real problem. This could be: 'We have to plan 10,000 products, but we do not have the resources to do that.' Then, this is a real problem. You can then check: 'How could we solve this with ML?' and: 'Do I have enough historic data to automate it?'" (i4)

Yet, the experts (88%) highlight that, when following a **(2) data-driven ideation** instead, potential problems are first explored that build on available data. Then, organizations explore expected organizational value of its solution. If no significant purpose can be identified at this point, solving the problem should be dismissed or postponed due to its missing added value:

"We started with data. We accessed the data and investigated whether we can find things where we think that we can build something out of it. [...] If you do this without business, then this will not get you very far. You need domain knowledge to verify or falsify your hypotheses. Otherwise, it could be that you built something that works, but then they say: 'Thank you, but this does not help us at all.'" (i6)

3.4.2 Evaluation Phase

While the initial ideation phase aims to uncover potential problems, the experts (92%) further outline a subsequent evaluation phase that aims to assess the likeliness of a problem's particular nature and context being suited to be solved with ML-based AI:

"After letting our thoughts run freely for a little bit to explore potential cases, we then narrow them down at an early stage. We try to strongly intervene to ensure that it does not go into every direction as this may result in: 'It's a nice idea, but this is not really suitable for this approach.'" (i5)

Our interviews indicate that this can be achieved with two evaluations that allow assessing the likeliness of a problem's suitability to be solved with ML-based AI. Besides, the evaluation phase also yields first indications for basic design decisions regarding ML-based AI particularities (as we will discuss). As Figure 2 illustrates, the evaluation phase comprises two evaluations: With our interviews, we found that the evaluation phase usually starts with (1) *evaluating problem* *substance*; that is, the assessment of hard key factors that must be fulfilled to render a problem potentially solvable with ML-based AI. We further found that organizations then usually proceed with (2) *evaluating problem particularities*; that is, the assessment of rather soft factors that must be fulfilled in a certain degree special to a given problem to render AI solutions favorable and useful. Below, we detail the two identified assessments of the evaluation phase.

3.4.2.1 Evaluating Problem Substance

With our interviews, we identified key factors that the experts considered to form the substance that a problem must fulfill to be generally suited to be solved with ML-based AI. We refer to them as "substance" as they represent hard evaluation criteria that must be satisfied by each problem. Otherwise, a problem renders generally unsuited to be solved with AI. We identified five key factors that likely form the problem substance:

First, the factor that the experts (83%) most frequently highlighted is the **(a) indefinability of sufficient human-designed rules**. The experts stress that if it is possible and feasible for humans to derive a sufficient solution and translate it into a set of rules, then it is usually not favorable to solve the problem using ML-based AI, but to create a solution performed manually or by a non-ML IS instead. This is, because ML-based AI solutions entail potentially detrimental properties that can render them less useful and even impractical in certain contexts (see section 3.4.2.2.). The experts therefore emphasize that organizations should consider ML-based AI as a second-choice problem solver that becomes only employed for problems for which their human counterparts fail to derive or articulate sufficient solutions. For instance, this is the case if humans can only define rules that cover a problem partially or must spent extensive efforts to update defined rules over time:

> *"First, I always ask: 'Do we really need ML?' If I can define sufficient rules, then I would always use these to better guarantee correct solution behavior. Only if such human-defined solutions are insufficient or come with high maintenance efforts, I would try to solve it with ML." (i4)*

Second, a factor that the experts (71%) also highlight is a problem's **(b) self-containment** in terms of its framing. The experts stress that organizations must be able to make data available that describes the aspects that are key to a problem. That is, because ML-based AIs can only observe a problem through the data that they consume and cannot consider any non-captured aspects in their solutions. The experts further stress that a representative, self-contained problem framing does not only depend on an organization's data availability and quality, but also on the problem's nature itself. If a problem solution requires capabilities hardly capturable with statistics (e.g., intuition, long logic chains, or common sense) an ML-based AI will likely fail to find a reasonable solution within the data:

> *"The AI must be able to consider all information that I need to make a decision. [...] Sometimes, you want to use an AI for all steps, but you missed the multiple logical steps involved that can't be solved by a single AI." (i13)*

Third, the experts (71%) also frequently highlight that organizations must become aware that ML-based AIs only solve problems by deriving inferences through induction, i.e., deriving general, statistical patterns from specific observations to solve a problem. The experts thus stress that organizations must be able to frame their problems in such a way that a **(c) generalization** is aimed as solution. They further emphasize that organizations should deliberately evaluate whether they can expect any derivable patterns that might be integrated into a general problem solution. Otherwise, it becomes likely that their ML-based AI cannot derive a sufficient solution:

> *"Many projects fail because you cannot generalize related aspects well enough. It's just like that. You should carefully assess upfront whether processes, that you expect to follow a certain structure, are really likely to do it." (i6)*

Fourth, another factor stated by the experts (58%) is that organizations must be able to define a **(d) clear, measurable goal** of a problem to allow for an ML-based AI solution. Due to the usually large solution space that ML-based AIs explore, the experts stress the necessity of clearly defining a goal and related metrics. Otherwise, an ML-based AI will not be able to distinguish between good and bad outcomes, essentially stopping it from producing an optimized solution. The experts thus highlight that problems for which an organization's members cannot agree on the correctness of potential outcomes and how to measure the outcome's quality are not suited. Especially in subjective problem domains (e.g., rating beauty; Rhue, 2019), the experts stress that an agreement on what the organization perceives to be correct is inevitable:

> *"You must precisely define the problem: 'What exactly is the goal you're trying to achieve?' and to define an evaluation metric for that goal to grasp and agree on how a sufficient solution would look like." (i10)*

Finally, a fifth key factor that the experts (54%) emphasized is that organizations must be able to ensure a **(e) continuous adaptation** of an ML-based AI's solution. In particular, they stress that organizations must ensure a frequent monitoring, data collection, and retraining on more current or comprehensive data bases to let ML-based AI solutions evolve over time. Otherwise, the organizations cannot ensure well-performing solutions if the conditions that affect related problems change or relevant exceptions are mistreated by the solutions. The experts stress that this adaptation represents a necessity for keeping the solution useful, emphasizing that organizations must evaluate whether the context allows for continuous monitoring and revision of ML-based AI solutions:

> *"I must be able to evaluate the results regularly to see if the solution still makes sense. Just because nobody checked if the data or business processes had changed, one of our well-working AIs rendered completely useless over time." (i4)*

3.4.2.2 Evaluating Problem Particularities

While the factors that comprise problem substance represent hard factors required to be fulfilled, the experts also described rather soft factors that might be acceptable in various degrees as long as a problem-specific minimum can be ensured. We refer to them as "problem particularities" as the required form of such factors is specific to a problem's particular context. Based on our interviews, we identified five factors that emerged to likely form key problem particularities. We found that such factors are usually evaluated subsequent to confirming a problem's substance.

First, as an ML-based AI solution is based on statistical generalizations, it will certainly produce errors at some point (Brynjolfsson & Mitchell, 2017). Thus, the experts (92%) stress that the degree of **(a) tolerable erroneousness** for each problem context has to be evaluated. In particular, organizations should explore potential error types along with their degree of severity. The organizations should then clarify each type's maximum tolerable rate that must be ensured by a solution. The experts also emphasize that organizations must understand whether the absolute avoidance of certain errors is vital for a problem's solution to avoid a detrimental ML-based AI solution. This can also be the case if the maximum tolerable rate is not expected to be achievable or no reasonable mechanisms to intercept such errors can be identified (e.g., humans revising AIs' outputs):

> *"ML will always make some mistakes. So I have to ask myself: Can I allow errors to occur? How much worse is a false negative compared to a false positive and with which error rate am I willing to live? [...] Which protective layers can I build around it?" (i1)*

Second, the experts (83%) state that organizations should evaluate the relevance of **(b) transparency** of the inner workings and reliability of ML-based AI solutions. As the achievable transparency level varies across algorithms (Diakopoulos, 2016), the experts stress that organizations must understand their transparency requirements for each specific problem context. Depending on the problem, a solution must also offer the possibility to explain why a result is being provided. The experts warn that if organizations ignore any transparency requirements, their absence may create distrust among users, potentially even resulting in a complete usage refusal. Providing transparency may also be required to meet regulatory requirements:

> *"The lack of understanding of what is happening in the AI might be critical as algorithms may pay attention to completely different things than what we think they pay attention to. However, if and how big a problem this is actually depends on the context of its intended use." (i21)*

Yet, some experts (38%) also emphasize that organizations often ask for ungrounded, high transparency levels, even if they are not required. The experts warn that asking for excessively high transparency may restrict or even hinder achievable solutions as this can limit the design of AI solutions. Organizations should therefore carefully explore given problem contexts to uncover actual transparency needs:

> *"It's interesting that AIs' transparency is seen so critical, as we've already given up understanding many things that happen in our world—hardly anyone knows how their refrigerator works. We can evaluate AIs statistically. For many contexts this is completely sufficient to know." (i6)*

Third, a further factor that the experts (67%) highlight is **(c) fairness** of AI solutions. The experts stress that organizations can use ML-based AI to actively promote more objectivity through reducing individuals' habits (e.g., prejudices or corruption) by generalizing over multiple individuals' behavior. Yet, the use of ML-based AI solutions may also create new ethical issues if AIs create discriminating behavior due to data being biased towards certain preferences and prejudices, or ethnical and social groups being badly represented in the data. The experts thus emphasize to carefully assess existent or potentially arising ethical issues within data and organizational contexts that may affect or be affected by ML-based AIs' fairness:

> *"Any systemically incorporated data bias will be adopted by AIs. For example, such AIs may discriminate customers that a bank's employees used to discriminate against traditionally. This needs special attention, but not everyone is aware of existing or potentially arising biases. So, we must take deliberate action to uncover discriminatory biases." (i9)*

Fourth, while manual solutions or IS that are based on human-designed rules can be usually easily adapted to fulfill specific requirements, the experts (46%) stress that solutions derived with ML-based AI may only offer a limited **(d) customizability**. Although it is possible to customize an AI solution by adding rules which modify its input or output, a fundamental change of the model-driven behavior can become problematic depending on the applied ML algorithm; that is, because conducting adaptations of more complex, non-transparent algorithms (e.g., neural networks) requires indirect changes through retraining AIs with a changed goal or data and thus knowledge of data-science techniques in combination with domain and data understanding. The experts therefore emphasize that organizations must carefully assess whether problems have to meet any potential requirements that might be too specific to be ensured at the core of an AI solution:

> *"Our AI recognized a billion documents correctly, but then we had one type that didn't want to work right. It can be an incredible effort to also correctly get this type while ensuring the correctness of the other cases. In traditional programming, I can just add a rule to handle this. But if I treat exceptions in my AI, I can't just handle that exception, I actually start to solve the problem all over again." (i17)*

Lastly, a fifth factor stated by the experts (42%) is the achievable **(e) response time** of an AI solution. Depending on the data volume that has to be processed and ML algorithms' processing time, the experts stress that a solution's response time may vary widely. As with human solutions however, organizations must ensure to provide a response as soon as it is required in the specific context. A gradual solution might not only render produced outcomes useless but may even cause fatal consequences (e.g., delayed warnings). Thus, organizations should carefully evaluate required minimum response times to clarify whether available data and algorithms likely allow for a fitting solution:

> *"If an organization wants to get a result every five minutes, then, of course, it has to be ready within five minutes. That's something you must always actively examine: 'How often do you need a result? What is the time horizon? Is it even realistic that we do the inference in time?' Depending on the ML algorithm you want to use, the inference can take a while." (i18)*

3.5 Discussion

With our study, we explored how problem finding translates to organizational contexts in which solutions are not purely derived by humans, but by ML-based AIs. Through interviewing 24 experts that regularly conduct AI initiatives, we found first evidence for essential procedural artifacts and related key factors. We synthesized our findings to propose a basic problem-finding framework that is contextualized for ML-based AI problem solving.

We can offer several theoretical contributions. To the best of our knowledge, we are among the first to study how the essential problem finding translates to organizational ML-based AI solver contexts. By providing initial findings on how to identify problems and evaluate their suitability to be solved with ML-based AI, we answer recent calls for research on how to manage the emerging human-machine symbioses in cognitive organizational contexts (e.g., Coombs et al., 2020; Rai et al., 2019; Rzepka & Berger, 2018; Schuetz & Venkatesh, 2020). Moreover, our findings qualify the recently emerging lines of IS research that examine ML-based AI particularities, such as research on fair (e.g., Afrashteh et al., 2020; Martin, 2019; Rhue, 2019) and explainable AI (e.g., Diakopoulos, 2016; T. Miller, 2019; Peters et al., 2020), and confirm the relevance of their consideration even at the earliest project stage of problem finding. Our findings further confirm that the fundamental ideation and evaluation phases of purely human-driven problem finding also form key phases in AI solver contexts. Lastly, as we identified key characteristics of AI-suited problems, we hope to inform future research on innovation, design, and diffusion of ML-based AI in organizations.

Our study also comprises significant contributions for practitioners. Organizations can use our findings to better manage their problem-solving activities when their humans and machines jointly contribute to problem solving. In particular, organizations can apply the proposed framework to explore and evaluate problems underlying potential AI initiatives in a structured manner. Our findings can help organizations to better uncover possibilities to exploit ML-based AIs' potential for enhancing processes and offerings. Organizations can also use the framework to protect themselves already at an early stage from mistakenly promoting AI-driven problem-solving initiatives that are not suited for deploying ML-based AI.

Of course, our study is subject to some limitations. First, we did not perform empirical testing of the proposed framework. Here, future studies should focus on evaluating both the procedural artifacts and identified key factors. As we chose to pursue a rather general perspective, future studies can test our findings' applicability in contexts with special requirements to further contextualize the model (e.g., highly serious or subjective contexts, such as medical or recruiting solutions). Second, while we tried to ensure a wide-ranging analysis, the resulting set of proposed factors is rather a non-exhaustive list. As we focused on exploring key artifacts and factors, further analyses that offer additional explorable contexts and problem characteristics represent fruitful avenues for future research. Third, although our interviewees cover a wide range of roles and industries, we cannot completely rule out any data biases. To reveal such biases, quantitative studies in varying contexts can be used to further validate our framework's applicability.

Presently, we do not know much about how ML-based AI will change organizational problem solving. Yet, history showed us that ML-based AI is able to build brilliant solutions if applied to the right problems. Furthering our understanding of how to effectively uncover suitable problems may therefore play a crucial role in ultimately unlocking the full potential of AI.

3.6 Acknowledgements

This study has been funded by the German Federal and Hessen State Ministry for Education and Research within their support of the National Research Center for Applied Cybersecurity ATHENE.

4 Paper B: Artificial Intelligence and Individual-level Performance (Focus: Task–Technology Fit)

Title

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

Authors

Sturm, Timo and Peters, Felix

Publication Outlet

Proceedings of the 28th European Conference on Information Systems (ECIS 2020)

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 patterns 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., 2015), 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" (Legg & Hutter, 2007; Nilsson, 1998; Poole et al., 1998; Russell & Norvig, 2021), which is "anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators" (Russell & Norvig, 2021, p. 34). It defines intelligent behavior as an agent function that selects executable actions based on current context information (Russell & Norvig, 2021). This function can be realized with various approaches, e.g., manually defined rules or statistics (Russell & Norvig, 2021). The approach that enabled recent AI advances (e.g., He et al., 2015; 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; Mitchell, 1997). Without ML, solutions are codified entirely by humans, e.g., humans designing rules to define robots' routines (Russell & Norvig, 2021). 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., 2017), 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 3 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.

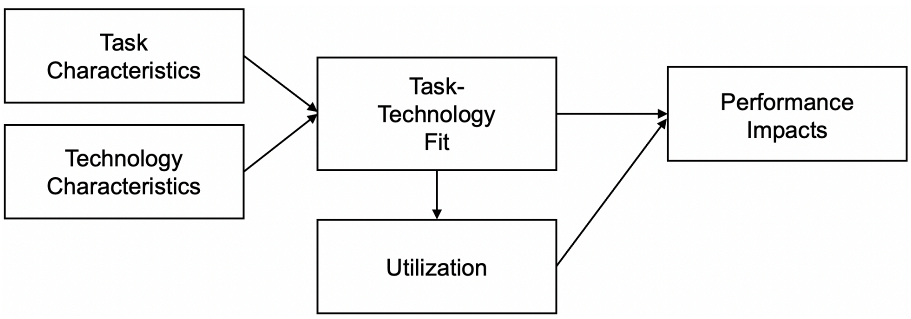

Figure 3: The TTF Theory as Conceptual Base (Goodhue & Thompson, 1995)

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 (e.g., 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. (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, Wongpuninwatana 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 organization 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 & 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 & 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., 2013) 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 3 provides detailed information on the involved experts.

Table 3: Experts who participated in the study

ID	Age	Gender	AI Experience	Profession	Firm
i1	51	Male	6 years in 2 industries	Data scientist	Provider firm
i2	39	Male	6 years in 2 industries	Data scientist	Provider firm
i3	33	Male	7 years in 5 industries	Manager	Provider firm
i4	31	Female	7 years in 5 industries	Presales consultant	Provider firm
i5	44	Male	3 years in 1 industry	Developer	User firm
i6	28	Male	3 years in 6 industries	Presales consultant	Provider firm
i7	30	Female	3.5 years in 4 industries	Technical consultant	Provider firm
i8	33	Male	6 years in 10 industries	Manager	Provider firm
i9	35	Male	5 years in 3 industries	Technical consultant	Provider firm
i10	32	Male	9 years in 4 industries	Data scientist	User firm
i11	25	Female	3 years in 1 industry	Manager	User firm
i12	34	Male	6 years in 3 industries	Data scientist	User firm
i13	32	Male	7.5 years in 1 industry	Data scientist	User firm
i14	32	Male	10 years in 4 industries	Technical consultant	Provider firm
i15	52	Female	4 years in 3 industries	Data scientist	Provider firm
i16	32	Male	10 years in 7 industries	Manager	Provider firm
i17	30	Male	5 years in 6 industries	Presales consultant	Provider firm
i18	37	Male	5 years in 3 industries	Data scientist	Provider firm
i19	36	Male	8 years in 1 industry	Manager	User firm
i20	35	Male	12 years in 2 industries	Data scientist	Provider firm
i21	37	Male	7.5 years in 3 industries	Data scientist	Provider firm
i22	41	Male	3 years in 1 industry	Developer	User firm
	39	Male	3 years in 1 industry	Developer	User firm
i23	35	Male	6 years in 4 industries	Technical consultant	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 (Agrawal et al., 2017; Brynjolfsson & Mitchell, 2017; Fedyk, 2016; Satell, 2018), to compare our findings with existent knowledge (see section 4.4), 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 4.

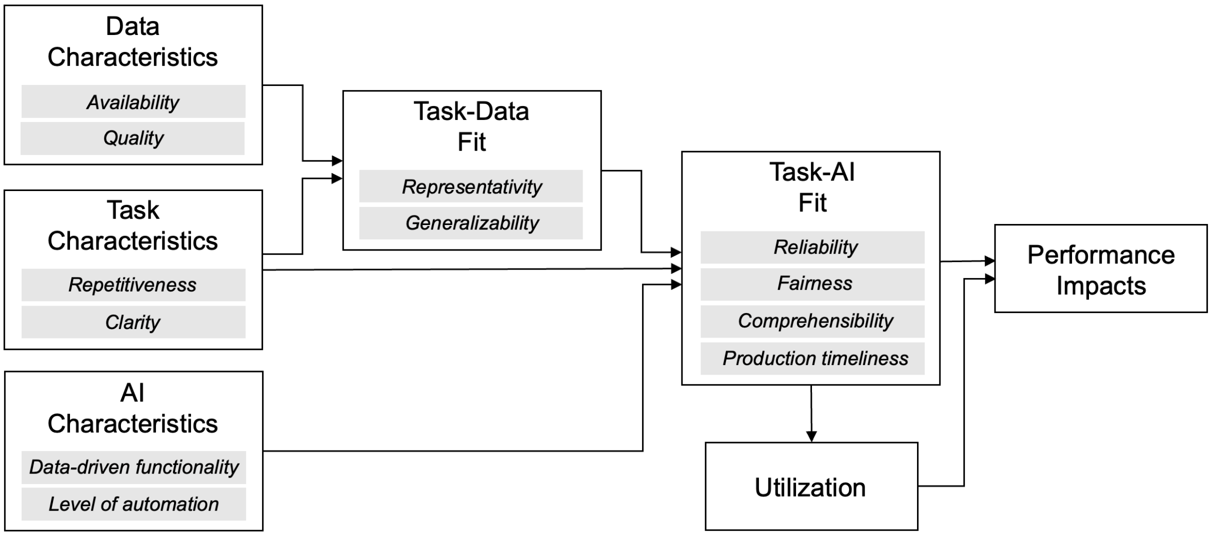

Figure 4: Extended and contextualized theoretical model of TTF in the context of ML-based AI (due to brevity, we abbreviate "ML-based AI" with "AI" in the construct names)

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 task-data fit to impact task-AI fit besides task and AI characteristics. Lastly, we contextualized "task-technology 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 (Agrawal et al., 2017; Fedyk, 2016; Satell, 2018). 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 (Agrawal et al., 2017; Clarke, 2016; Ghasemaghaei et al., 2018). 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 (Agrawal et al., 2017; Brynjolfsson & Mitchell, 2017; Traumer et al., 2017). 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 & 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 (Agrawal et al., 2017; Brynjolfsson & Mitchell, 2017). 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 (Brynjolfsson & Mitchell, 2017; Fedyk, 2016; 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 non-generalizable 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 (Bishop, 2006; Samuel, 1959). 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 & 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 (Agrawal et al., 2017; Traumer et al., 2017). 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 & 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., 2017). 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 & Mitchell, 2017; Guidotti et al., 2018; T. 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 & 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 & Norvig, 2021). While complex algorithms may result in higher predictive performance (Kaplan et al., 2020), their complex operating principles may reduce their comprehensibility (T. 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 trade-offs 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 & Berger, 2018) and ethical considerations, such as transparency (e.g., Chai & Li, 2019; Fernandez et al., 2019) or fairness (e.g., Haas, 2019; van den Broek et al., 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 organization-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 & 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.

4.6 Acknowledgements

This study has been funded by the German Federal and Hessen State Ministry for Education and Research within their support of the National Research Center for Applied Cybersecurity ATHENE.

5 Paper C: Artificial Intelligence and Group-level Performance (Focus: Organizational Learning)

Title

The Case of Human-Machine Trading as Bilateral Organizational Learning

Authors

Sturm, Timo; Koppe, Timo; Scholz, Yven; and Buxmann, Peter

Publication Outlet

Proceedings of the 42nd International Conference on Information Systems (ICIS 2021)

Abstract

In today's organizations, both humans and machine learning (ML) systems jointly form routines. Yet, we do not know much about the underlying reciprocal interplay between them, which complicates their effective coordination. Taking an organizational learning perspective, we study the dynamics of human learning and ML to understand how organizations can benefit from their respective idiosyncrasies when enabling bilateral learning. Drawing on a case of human traders and a reinforcement ML system trading productively at Allianz Global Investors, we apply human-machine pattern recognition on digital trace data to explore their (interconnected) dynamics. We find that bilateral learning can increase trading performance, which appears to result from an emerging virtuous cycle between humans and the ML system. Our explorative case study offers insights into how organizational learning depends on the coordination of both human learning and ML, which can help manage the collaboration between human and artificial intelligence within organizational routines.

Keywords

Machine learning, organizational learning, digital trace, human-AI collaboration

5.1 Introduction

In recent years, numerous breakthroughs in artificial intelligence (AI) have uncovered AI's potential to surpass human performance in various contexts (e.g., He et al., 2015; Vinyals et al., 2019). In the light of such breakthroughs, more and more organizations strive to use AI in their processes to improve their organizational performance (e.g., Bean, 2019; Forbes Insights, 2018). In doing so, today's organizations focus on using AI to automate (sub-)tasks within routines (e.g., Brynjolfsson & Mitchell, 2017). To date, it is undeniable that the rise of such AI-enabled automation already transforms organizations' routines, especially when AI takes over tasks that were formerly performed by humans (e.g., Raisch & Krakowski, 2021). In this regard, existent discussions mainly deal with achievable cost savings and error reduction with AI-enabled automation (e.g., Kellogg et al., 2020), shifting humans to other 'higher-value' roles (e.g., Brynjolfsson & Mitchell, 2017), or emerging social challenges such as ethical AI (e.g., Rhue, 2019). Only recently, discussions began to stress the great importance of the reciprocal interplay between humans and intelligent machines for their coordination and its consequences within organizations (e.g., K. Leavitt et al., 2021; Murray et al., 2021; Rai et al., 2019; Schuetz & Venkatesh, 2020). While a few researchers have already begun to examine how human actions affect AI and vice versa and how organizations may coordinate this relationship, related research still remains in its infancy and emphasizes the need for further analyses (e.g., Grønsund & Aanestad, 2020; Lyytinen et al., 2021; Seidel et al., 2019; Sturm, Gerlach, et al., 2021). Surprisingly, one aspect has received little attention, even though it is central not only to the technology behind modern AI, but also to its relationship with humans and organizational routines: *learning*.

The technology that enables modern AI is machine learning (ML). AI systems based on ML—by us referred to as *ML systems*—use ML algorithms to derive patterns from data and then apply these patterns to new data in order to act (Mitchell, 1997; Russell & Norvig, 2021). By doing so, ML systems do not require us to manually solve and translate our solutions into code anymore but derive solutions on their own from given data (Samuel, 1959). In other words, ML algorithms grant information systems (IS) the ability to learn autonomously to act intelligently (Brynjolfsson & Mitchell, 2017). With their ability to learn, ML systems join the central process of *organizational learning* beside human learners (e.g., Argote et al., 2021; Ransbotham et al., 2020). Organizational learning is the fundamental driver that controls how strongly an organization relies on and adapts established routines and how strongly it adopts new ones (March, 1991). Organizational learning thus controls how an organization adapts itself to its environment and, by doing so, defines an organization's performance (e.g., Argote & Miron-Spektor, 2011). Organizational learning is based on a complex system of learners that interact with one another, which requires coordination (March, 1991). Due to its high complexity, optimal coordination of organizational learning constitutes a difficult endeavor (e.g., Levitt & March, 1988). Despite decades of research, however, literature has largely assumed the learner to be purely *human* (e.g., Argote et al., 2021). With ML being able to learn as well while differing significantly from human learning (as we will discuss), the rise of ML denies this assumption and requires us to fundamentally rethink organizational learning theory. Yet, we only know little about how ML systems actually affect organizational learning (Argote et al., 2021). As the result of organizational learning is more than only the sum of individual learning but also subsists of the individuals' interactions (Argote et al., 2021; March, 1991), mutual learning that builds on humans' and ML systems' individual learning should not be neglected. Organizational learning thus constitutes a promising context for analyzing the bilateral relationship between humans and learning machines that collaborate within their organization (Sturm, Gerlach, et al., 2021).

To help unravel the complex bilateral human-machine relationship in organizational learning, we conducted a case study at Allianz Global Investors, a global asset management firm that introduced an autonomous ML system for trading financial instruments next to its human traders. Trading constitutes a fruitful context to study organizational learning as learning lies at the heart of trading: organizations aim to learn about the complex causal structure of markets and related trading strategies in order to optimize their future investment endeavors. Moreover, such trading is executed in a purely digital world in which market states and trading actions are naturally tracked and are rich in information. By exploring digital trace data, we thus aim to answer the following research questions (RQs): *In the context of trading, (1) how does ML and human learning affect each other in organizational learning, and (2) how can organizations leverage their bilateral relationship to improve organizational performance?*

To answer the RQs, we rely on an abductive, pragmatist approach for human-machine pattern recognition to analyze the (interconnected) dynamics of the humans' and ML system's trading behavior. In doing so, we explore how idiosyncrasies of human learning and ML contribute differently to trading and how their synthesis affects the organization's trading performance. Our case study offers empirical insights about how organizational learning depends on the coordination of both human learning and ML, which can help organizations to craft effective human-AI collaboration designs and stimulate related research endeavors.

5.2 Theoretical Background

We first introduce organizational learning and ML along related work. Next, we compare idiosyncrasies of human learning and ML. We then combine both research streams to form our study's objective.

5.2.1 Organizational Learning

In their seminal work, Levitt and March view organizational learning "as learning by encoding inferences from history into routines that guide behavior" (Levitt & March, 1988, p. 320). Individuals in organizations gather experiences based on their chosen actions and associated outcomes. Based on these experience-outcome pairs, they infer learnings (i.e., beliefs about the causal structure of reality) to guide future actions. Organizations store these learnings in their routines to make use of and further distribute the developed knowledge (e.g., Argote & Miron-Spektor, 2011). By doing so, organizations form complex systems of interacting individuals who learn to make sense of the environment in which organizations act and to which they adapt to (Levitt & March, 1988). The better organizations learn (i.e., the better they understand their environment to guide their actions), the better they can act and adapt to increase organizational performance (March, 1991). Organizational learning marks therefore an essential process that organizations need to pursue continually in order to survive (e.g., Grant, 1996; March, 1991).

One of the most central and crucial concepts in organizational learning is that organizations need to balance explorative and exploitative learning (e.g., Gupta et al., 2006; March, 1991). While explorative learning represents the search for new ideas with uncertain outcomes which shift away from an organization's current knowledge, exploitative learning refers to the use and incremental refinement of existing knowledge to obtain its immediate benefits (March, 1991).

Balancing both types of learning is important: If organizations overemphasize the short-term benefits they can gain by exploiting given knowledge, they can become trapped in a state of stagnation, ignoring new potentially useful directions. In contrast, if organizations neglect exploitation while extensively exploring new ideas, they will not survive in a competitive environment as they fail to refine and apply knowledge to develop specific competences (March, 1991). Achieving this balance is, however, a difficult endeavor that has yielded decades of research studying how to overcome the various flaws of organizational learning to optimize organizational performance (e.g., Argote & Miron-Spektor, 2011; Levitt & March, 1988). Especially the so-called 'learning myopia', which is the tendency to favor exploitation over exploration, represents a major and versatile issue: Due to their distant and uncertain benefits, learners tend to avoid experimenting with new, yet unproven ideas and prefer to rely on established ideas that proved to be successful in the past (e.g., Levinthal & March, 1993). Research has uncovered numerous factors known to either mitigate (e.g., high team diversity; March, 1991) or intensify (e.g., incentives that reward successes and penalize failures; March, 2010) a learner's myopia, further complicating the crucial balance of explorative and exploitative activities (Levinthal & March, 1993). Despite decades of research on organizational learning (several fantastic overviews exist, e.g., Argote et al., 2021; Argote & Miron-Spektor, 2011; Huber, 1991), research on the impact of IS on organizational learning still remains in its infancy (Argote et al., 2021; Argote & Miron-Spektor, 2011).

5.2.2 Machine Learning

One of the most widely accepted concepts of AI is the one of the 'intelligent agents', which is "anything that can be viewed as perceiving its environment through sensors and acting upon that environment" (Russell & Norvig, 2021, p. 34). Here, intelligent behavior is defined as an agent function selecting actions based on context information. While various approaches exist to realize this function (e.g., human-defined rules or statistics; Russell & Norvig, 2021), the one that largely underlies modern AI is ML; that is, learning with algorithms from data-based experience to infer models that capture derived data patterns. ML systems then apply these models to new data to guide their behavior (Jordan & Mitchell, 2015; Mitchell, 1997). ML systems are developed in an iterative process, in which humans select and prepare data, select and parametrize algorithms, and assess implemented alternatives to craft the best-performing ML system. By doing so, humans define the conditions under which the ML system learns to develop its own understanding of a problem solution (e.g., Amershi et al., 2019; Sturm, Gerlach, et al., 2021). So far, IS research on ML has mainly focused on topics like adoption (e.g., Pumplun et al., 2019), automation of (sub-)tasks (e.g., Brynjolfsson & Mitchell, 2017), or emerging social challenges like ethical or transparent AI (e.g., Rhue, 2019). Only recently, scholars began to stress the great importance of the reciprocal interplay between humans and AI (e.g., K. Leavitt et al., 2021; Murray et al., 2021; Rai et al., 2019; Schuetz & Venkatesh, 2020). While a handful of researchers have begun to examine how humans affect AI and vice versa and how organizations may coordinate this relationship, related research remains scarce and emphasizes the need for further analyses (e.g., Grønsund & Aanestad, 2020; Kellogg et al., 2020; Lindebaum et al., 2020; Lyytinen et al., 2021; Murray et al., 2021; Seidel et al., 2019; Sturm, Gerlach, et al., 2021). Especially, work on the impact of ML on organizational learning is still limited and mainly provide only basic insights into potential setups and hypothetical consequences (e.g., Ransbotham et al. (2020) propose possible learning modes with varying human-machine autonomy, Balasubramanian et al. (2022) theorize and simulate potential consequences of ML).

In doing so, the scholars primarily put forth one foundational, ongoing discussion: does ML amplify (e.g., Balasubramanian et al., 2022) or alleviate (e.g., Sturm, Gerlach, et al., 2021) learning myopia, shifting organizations towards exploitation or exploration?

5.2.3 Strengths and Weaknesses of Human and Machine Learning

Organizational learning is generally far from perfect (e.g., Argote et al., 2021; March, 2010). Reality itself complicates learning as its complexity renders its underlying causal structure difficult to discern: numerous variables interact and can change continuously, including actual random variations (March, 2010). In an ideal world, a learner thus gathers extensive experiences (i.e., samples of reality) to construct a complete picture of reality and flawlessly sees through complex multivariate relations while ignoring random noise to infer accurate learnings about reality (March, 2010). To assume such ideal circumstances is, however, rather lunatic than appropriate as learning is constrained by humans' limited experiences and learning capabilities (e.g., Levinthal & March, 1993; Simon, 1972). As ML systems differ significantly in their way of learning, often even viewed as a panacea to overcome human limits (e.g., Lindebaum et al., 2020), we now compare major idiosyncrasies of humans and ML systems. To this end, we focus the comparison on key elements of learning that are essentially shaped by a learner's capabilities (e.g., Argote et al., 2021; Levinthal & March, 1993; Levitt & March, 1988): (1) *learning base,* (2) *inference*, and (3) *learned model.*

a. Learning Base (Human vs. Purely Data-based Experience). A human observes reality through her/his unique sample of experience (Levitt & March, 1988). This sample is gathered by the human choosing and performing actions from among action alternatives (e.g., making a specific decision) and observing associated action outcomes (e.g., the perceived success or failure of a performed decision) over the human's lifetime (Argote & Miron-Spektor, 2011). To this end, human experience is far from being an optimal base for learning: our experiences usually comprise very small, incomplete samples of reality that are often skewed and erroneous (March, 2010). That is, because a single one of us cannot observe the overwhelming breadth of reality in its entirety but is limited to her/his specific interests, social and organizational context, unique sequence of decisions, repeatability of actions and contexts, measurement errors and misinterpretations, cognitive memory, and attention—just to name a few factors (e.g., Argote et al., 2021; March, 1991). A more comprehensive and correct picture of reality is therefore likely spread across multiple humans' diverse experiences (March, 2010). Humans thus usually take part in a time-consuming social learning process to jointly share, evaluate, and combine individual experiences to some extent (e.g., March, 1991). Yet, not everything is bad about human experience as a basis for learning. Human experience is not limited to specific media or domains per se. Humans can draw upon a rich amount of diverse integrable sources (e.g., knowledge repositories or human senses) to form their experiences and can transfer learnings and experiences between domains and contexts (e.g., Argote et al., 2021). Moreover, humans are able to contextualize learnings and craft hypothetical samples (i.e., thinking about 'what-if' scenarios), allowing them to enrich their small samples of reality (March et al., 1991).

In contrast, ML systems purely learn from data (e.g., Brynjolfsson & Mitchell, 2017; Mitchell, 1997). Indeed, organizations' data is also often skewed, erroneous, and incomplete, and is therefore often also far from being a perfect learning base. Yet, while organizations can only partly control and improve individuals' collected experiences used for learning (e.g., Argote & Miron-Spektor, 2011), ML systems' data is usually actively assessed, enlarged, and cleansed, partly mitigating the issues related to the less controllable human experiences (e.g., Amershi et al., 2019; Domingos, 2012). As an ML system can store and process large amounts of data, it can learn from a more comprehensive and diverse sample of reality, often covering multiple individuals' experiences (e.g., Brynjolfsson & Mitchell, 2017; Jordan & Mitchell, 2015). ML systems can therefore grant organizations the ability to craft and learn from larger and actively cleaned samples of reality. Yet, this can only be true for experiences described by ML systems' narrow frame (i.e., predefined goals and provided data; Salovaara et al., 2019). While ML systems' data-driven learning allows them to analyze larger samples of experiences, they are at the same time blinded by it: ML systems neglect any information outside their data frame, potentially overemphasizing aspects captured by data while ignoring aspects that are actually relevant but non-capturable with data—ML systems are unable to look outside the box (Domingos, 2012; Salovaara et al., 2019). ML systems' learning can therefore even act restricting as many learnings must be contextualized and critically reflected before applying them naively (e.g., Raisch & Krakowski, 2021). Moreover, ML systems usually require large amounts of data to learn reliably (Brynjolfsson & Mitchell, 2017; Salovaara et al., 2019). If only small data samples are available, no learnings can be derived or the ones available may rather confuse than benefit others who learn from resulting ML models (e.g., Balasubramanian et al., 2022).

b. Inference (Bounded vs. Formal Rationality). In addition to human experience being a non-ideal basis for learning, humans themselves are no perfect learners either that always flawlessly derive the causal structure from experience (e.g., Levinthal & March, 1993; March, 2010). That is, as famously coined by Simon, because humans can only learn within their bounded rationality: "boundedly rational agents experience limits in formulating and solving complex problems and in processing (receiving, storing, retrieving, transmitting) information" (Williamson, 1981, p. 553, quoting Simon). Humans simply struggle to untangle the complex relations of reality's numerous variables. Overwhelmed by reality's complexity, humans thus fall back on learning simplified heuristics to describe reality's causal structure instead of using complete optimization methods to derive an optimal representation of reality (March, 2010). Humans are also no rapid learners. Humans are slow in processing large amounts of information and take part in the slow social learning process to enrich their own learnings with the ones of others (e.g., Levinthal & March, 1993; March, 1981). Human knowledge creation certainly takes its time, which further impedes the creation of sound learnings if time is limited (March, 2010). Moreover, bounded rationality creates room for humans' irrational 'foolish' behavior (March, 2006). While acting foolish (i.e., not doing the seemingly 'right' things; acting imprudent or playful) is largely detrimental to organizations as it mostly yields costly failures (March, 2006), a small amount of foolishness can yet benefit organizational learning: Foolishness acts as driver for (unintended) exploration. Acting foolishly implies disregarding established beliefs about how things should be done, which often leads to trying out new (sometimes better) ways that would otherwise be overlooked, diversifying gathered experiences (March, 2006, 2010).

In contrast, rationality is imperative to ML systems: ML systems are deliberately implemented as rational agents with the explicit goal to always act (and learn) rationally (Russell & Norvig, 2021). Due to their high information processing capabilities and use of formal learning mechanisms, today's ML systems are even viewed as "supercarriers of formal rationality" (Lindebaum et al., 2020, p. 248), yielding hopes that organizations have finally crafted the perfect learners they have always hoped for (Lindebaum et al., 2020; Murray et al., 2021). As ML systems also rely on (indeed more profound) heuristics, they are, however, no perfect learners either—but indeed less bounded in their rationality: ML systems can analyze larger samples of experience and identify more complex relations between greater numbers of variables to derive more accurate heuristics than humans (e.g., Lindebaum et al., 2020; Raisch & Krakowski, 2021).

ML systems are also very efficient learners as they can process large amounts of data very quickly and can thus make new knowledge rapidly available as soon as new data exists (e.g., Kellogg et al., 2020; Lindebaum et al., 2020). Yet, having such increased rationality, ML systems may also alleviate foolishness. While this can be beneficial, ML systems also risk eliminating foolishness as an important mechanism to occasionally explore unorthodox ideas, which drive innovation (e.g., Balasubramanian et al., 2022).

c. Learned Model (Broad vs. Narrow). Humans learn mental models of reality (i.e., an individual's understanding of the world; Levitt & March, 1988). Such models generally cross multiple domains and contexts (e.g., Argote et al., 2021). For instance, a single human can learn a model on how to play an instrument, speak a language, and diagnose diseases. Humans can use their models to transfer learnings from one domain or context to another, which enables them to assess whether existing learnings likely fit novel or changed contexts (e.g., Raisch & Krakowski, 2021). In contrast to artificial general intelligence that aims to resemble the general focus of human intelligence, today's ML systems only enable narrow AI (e.g., Brynjolfsson & Mitchell, 2017; Sturm, Gerlach, et al., 2021): ML models are highly contextual models that purely focus on a narrow aspect of reality. While ML systems can adapt autonomously to changing contexts if the underlying concept does not change fundamentally, disruptive context changes or reduced information (e.g., through concept drifts, broken sensors) may confuse ML, leading to obsolete ML models that must be reevaluated and retrained by human experts that can look beyond its narrow frame (e.g., Lindebaum et al., 2020; Raisch & Krakowski, 2021). In other words, due to their high contextuality, ML models thus tend to be less robust to contextual changes compared to humans' mental models.

5.2.4 The Need to Revisit Learning During the Rise of Machine Learning

Organizations already use ML systems next to their human employees to autonomously shape, perform, and collaborate in organizational routines (e.g., Brynjolfsson & Mitchell, 2017). Yet, it remains unclear how humans and ML systems affect each other and how they should be coordinated as a whole. This is important: if done wrong, organizations may jeopardize organizational performance—and in the worst case their long-term survival if ML obstructs essential organizational processes (e.g., Raisch & Krakowski, 2021; Sturm, Gerlach, et al., 2021) by, e.g., exacerbating learning myopia (Balasubramanian et al., 2022) or spreading false beliefs (Sturm, Gerlach, et al., 2021). Only recently, management and IS scholars recognized the great relevance of managing the reciprocal interplay between humans and intelligent machines in organizations (e.g., Rai et al., 2019; Schuetz & Venkatesh, 2020). However, *learning*, although being a key aspect that blends human and ML systems' behavior, remains widely overlooked. This is surprising as focusing on learning allows to draw the discussions on human-machine collaboration back to its central driver and ML systems' specificity. Despite decades of research on organizational learning (e.g., Argote et al., 2021), the literature can only partly inform related studies as it has essentially assumed the learners to be purely *human*. Indeed, a few scholars already started to study ML's impact on organizational learning, but have still only scratched the surface and call for further research (i.e., Afiouni-Monla, 2019; Argote et al., 2021; Balasubramanian et al., 2022; Lyytinen et al., 2021; Ransbotham et al., 2020; Seidel et al., 2019; Sturm, Gerlach, et al., 2021). These studies are mainly theoretical work (exceptions: Ransbotham et al., 2020; Seidel et al., 2019), missing empirical insights and strongly suggest investigating human-machine learning empirically to enrich ongoing discussions. Although literature on ML in organizational learning is very limited, students of this topic can draw on widely established computer science and organizational learning literature to characterize human learning and ML (see previous sections). By doing so, focusing on learning allows us to ground analyses on the micro level (i.e., how human learning and ML function and differ) to contribute novel theory on the macro level (i.e., theorize human-machine dynamics). To this end, to provide empirical evidence, we now use the above-discussed characteristics of human learning and ML as a theoretical basis to explore 'macro-level' dynamics between humans and an ML system by studying a real-life case of human-machine trading.

5.3 Research Design

Below, we first introduce our trading case as a suitable empirical context to analyze organizational learning. Then, we outline our research approach and provide an overview of the collected data and its analysis.

5.3.1 Empirical Context

In our study, we examine the case of human-machine trading at Allianz Global Investors (AllGI), a global asset management firm with over 500 billion euros in assets under management in 2021 and more than two thousand employees worldwide. AllGI's investment routine follows two essential steps: (1) a portfolio manager requests an order (i.e., to buy a set of specific securities), and then (2) a trader executes this order on a best-efforts basis. Every month, AllGI's traders execute transactions worth billions of euros. In our case study, we focus on the second part of AllGI's process; that is, the actual execution of a given order. More precisely, we focus on the trading of futures contracts.[7] To this end, effective trading is an adaptive process that requires a trader to develop an understanding (i.e., learn a model) of the market environment and associated trading strategies to react purposefully. This is challenging as trading takes place in an extremely noisy, complex, and turbulent world: Financial markets change continuously and many factors affect the markets' development, rendering market comprehension and strategy development a very challenging endeavor. To be effective, today's traders choose from a large set of *trading algorithms*[8] that reflect execution sequences predefined by external brokers. Moreover, trading takes place in a world that allows to clearly monitor and evaluate trading behavior with every trading action and related market state being captured in data. Based on an industry-standard benchmark, each trading decision is evaluated with an associated (positive or negative) trading performance. The case of AllGI's trading therefore represents a well-suited context to explore organizational learning empirically: First, learning lies at the heart of trading as its success depends on thoughtful trading decisions. Second, as the whole trading process is conducted digitally, drawing on this case enables us to observe the traders' experiences and learned propensities over time within naturally collected digital trace data—no matter whether the trader is human or an ML system.

[7] Futures are financial derivatives for a transaction of an asset at a predetermined price and date.

[8] Not to be confused with ML algorithms, trading algorithms are manually programmed buying/selling rules used to automatically manage the price, size, and timing of trades, thus executing a predefined trading strategy chosen by a trader (Kissel, 2013).

5.3.2 Research Approach and Collected Data

Methodology. We conducted a descriptive case study (Yin, 2011) based on big data analyses of digital trace data. Digital trace data captures individuals' actions within organizational routines, thus allowing to explore actual individual behavior within specific empirical contexts over time (e.g., Lindberg, 2020; Pentland et al., 2020). We followed Lindberg's (2020) pragmatist approach for analyzing digital traces to demonstrate how patterns emerge from the idiosyncrasies of agents' actions in a particular empirical context (e.g., Lindberg, 2020; Venturini & Latour, 2009). The pragmatist approach emphasizes that actions can only be understood with regard to their specific context and associated meanings behind the different actions (Burks, 1946; Lindberg, 2020). Although pragmatists stress the importance of causation, pragmatism is less concerned with deriving universal patterns, but focuses on portraying contextually efficacious practices (Farjoun et al., 2015; Lindberg, 2020). In its essence, pragmatism aims to evaluate an action's meaning in terms of its consequences, as Lindberg (2020, p. 93) puts it nicely: "It is hard to observe internal emotional or cognitive states, but it is possible to clearly observe actions and the consequences that such actions engender. Thus, when trying to understand how actors think about and interpret their worlds, it is necessary to also look at their actions and the consequences of those actions. [...] [T]he pragmatist approach posits that understanding causality is central to understanding meaning, since the meaning of an action (or utterance, i.e., a speech act) largely resides in its consequences". Considering this perspective, Lindberg (2020) proposed a method grounded in human-machine pattern recognition. Using abduction, the method does not start with the a priori formulation of hypotheses, but with the discovery of patterns from digital trace data: First, based on human or machine pattern recognition, inductive generalizations are derived from data. In our case, we rely on machine pattern recognition (i.e., computationally derived patterns, e.g., descriptive statistics and correlations, or other regularities, e.g., action categorizations; Lindberg, 2020). The inductive generalizations function as 'working hypotheses' that are then justified using human or machine pattern recognition to explain inductive observations. In our case, we rely on human pattern recognition to explain identified patterns and reflect on our findings using extant literature and qualitative insights. Resulting inferences are "viewed as 'reasonable inferences' tempered by theoretical experience and intimacy with the data under scrutiny" (Lindberg, 2020, p. 94) which do not have to be inductively or deductively true but presumptively follow from the analysis. By doing so, related studies embody the capacity of science to make new discoveries by contributing fruitful evidence and innovative ideas to inspire and create momentum for further research endeavors and discussions (Dougherty, 2016; Lindberg, 2020).

Data. We accompanied AllGI's journey towards its adoption of a productive, autonomously trading ML system since the initial idea in January 2019. From its first rollout in mid-2020 until the end of November 2020, the ML system's initial knowledge base was formed and tested. We then collected data on its official productive use from December 01, 2020, to April 16, 2021. During this period, AllGI enabled bilateral human-machine learning beginning February 19, 2021. As one of the authors was employed as a trader for the whole duration by AllGI and was actively involved in the ML system's implementation, we were able to continuously gather first-hand insights into the adoption process, talk directly to the traders, receive access to internal documents, and attend internal meetings. We had full access to over 50,000 logged order events (i.e., data points capturing assignments, order routing, executions) with over 200 data fields per log entry. These logs enabled us to detailly review each step for every order from the initial request until completion for both the human traders and ML system. To study the digital traces, we focused our analyses especially on the execution price, benchmark price, chosen trading algorithms, and information about the given market state per trade. As can be seen in Figure 5, the captured traces include several orders (*N*) traded by either human traders (D_0, D_1) or the ML system (D_2, D_3) before (D_0, D_3) or after (D_1, D_2) the enabled ML system's advice. Moreover, to better understand the portfolio of over 190 trading algorithms, we had the chance to review algorithm manuals and talk to algorithm providers personally. We further had access to 52 meeting notes from AllGI's weekly team meetings and 45 meeting notes from AllGI's algorithmic trading reviews with the providers. We also received live demonstrations of the trading workflow, software, and ML system, and were able to talk to the traders directly about abnormalities that we identified in the data to gather further context information. After the observation period, we provided an anonymous online survey with open-ended questions in which the traders further reflected on their work with the ML system.

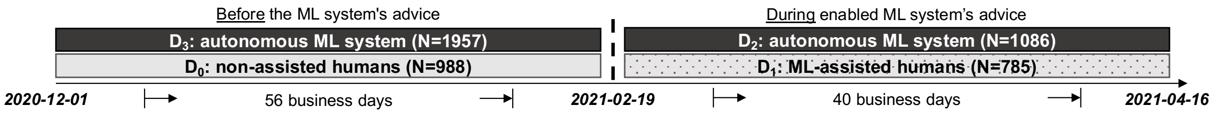

Figure 5: Collected Digital Trace Data

Analysis. We ensured research rigor by strictly following Lindberg's (2020) seven guidelines for crafting mutable digital traces and conducting abductive analysis; that is, we enriched our continuously sampled quantitative digital trace data captured by the trading system with additionally gathered qualitative data, iteratively solved puzzles that emerged from derived data patterns, searched for explanations for surprises, and investigated identified patterns' correlation and causation. Moreover, we ensured to satisfy Lindberg's (2020) principles for evaluating the process (i.e., developing theory) and product (i.e., the developed theory) of abductive inquiry to assure our findings' quality: First, to ensure a high-quality research *process*, the research problem, data, and analytical techniques must be well integrated. As highlighted before, we identified the trading context as a suitable context for analyzing the conundrum of (coupled) human learning and ML as learning lies at the heart of trading. Trading's digital nature allows us to track each actor's decisions, enabling us to collect quantitative data capturing individual trading decisions and context information that we further contextualized with qualitative data (e.g., from attending meetings and interviewing traders) to enable a somewhat 360-degree view on AllGI's trading process. We further ensured that the data pertains to the same individuals performing the same activities within the same organizational structures over time. Lastly, we also ensured the rigor of our analyses by enabling iterative cross-validation through continually exploring, confirming, and explaining patterns derived from quantitative data using machine pattern recognition in view of qualitative patterns deduced through human pattern recognition and vice versa. Second, to craft a high-quality research *product* with this method means to show "a process (consequences) that interacts with its environment (context), while at the same time also exhibiting the iterative dynamic between structure and agency (constitution)" (Lindberg, 2020, p. 103). We ensured this integration by analyzing the evolving (interrelated) dynamics of human traders and a trading ML system (i.e., constitution) in relation to varying market states (i.e., context) in terms of in- or decreasing trading performance (i.e., consequences) at all times. This allows us to assure that we situate identified practices and causal mechanisms within particular contexts to enable the identification of causal consequences of social structures. To further evaluate our results, we also judged derived findings against existent theory.

5.4 The Case of Human & Machine Trading as Organizational Learning

We first show how human-machine trading translates to organizational learning. Drawing on these insights, we then analyze how human, machine, and human-machine learning affects AllGI's trading performance.

5.4.1 Human Trading, ML-based Trading, and Bilateral Human-Machine Trading

Throughout the course of our study, AllGI traversed three different phases with each phase representing a unique learning scenario: (a) *purely human trading*, (b) *autonomous ML-based trading*, and (c) *bilateral human-machine trading*. Figure 6 illustrates a conceptualization of the human learning, ML, and their interconnections that can be observed in the three trading modes as discussed below.

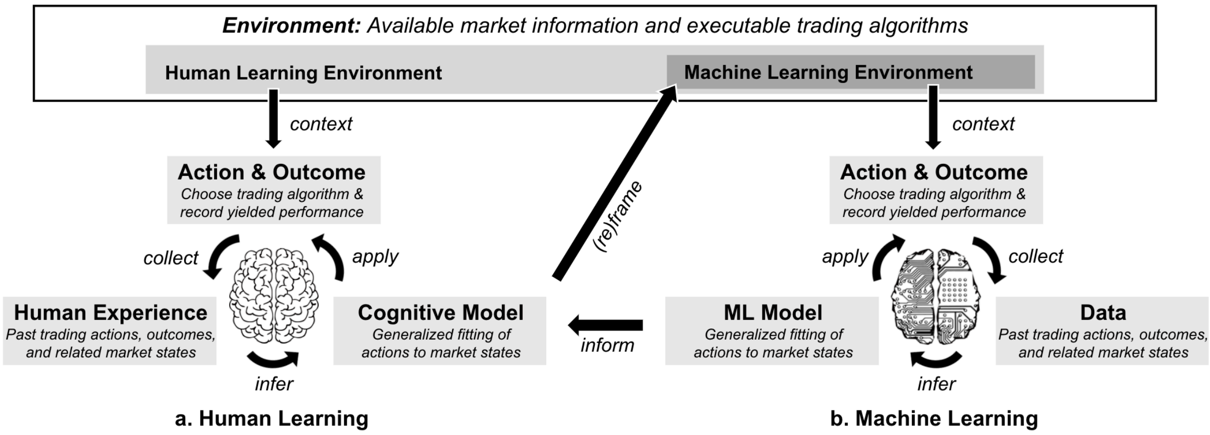

Figure 6: Conceptualized Bilateral Human-Machine Learning (in the Trading Case)

Illustrated in Figure 6a and Figure 6b respectively, both the human and ML system's learning cycle at AllGI follows the same essential logic: A learner's *action* involves comprehending the current market state to choose the seemingly best fitting trading algorithm from a set of executable algorithms. Each trading algorithm is externally defined and provided by external brokers. The learner then indicates her/his/its choice in AllGI's execution management system that applies the selected algorithm as a trading strategy to trade the given order. Hence, the action that is performed by a learner is the decision about a trading algorithm to fit the current market situation. Each trade yields some trading performance as an action *outcome*, defined as the margin between the trade's execution price and AllGI's performance benchmark.[9] The higher a trade's performance, the more successful a trade is regarded. Through executing trades, a learner samples action-outcome pairs and contextualizes these with gathered information about the faced market states. This sample represents a learner's *experience*. The learner then tries to derive generalizable patterns from this experience by inferring heuristics about the success and failure of each trading algorithm within specific market states. Based on these heuristics, the learner adapts her/his/its *model* that the learner uses to guide future trading actions in the quest to optimize trading performance. Building on this first-order individual learning cycle,

[9] The benchmark ('arrival price') is the market price at the time that the order starts working (i.e., arrives at the market). The benchmark gets adjusted if the order has a high impact on the market or the market moves shortly after the order started working. In those cases, the benchmark price is an average price over a time period instead of a single point in time. The benchmark is normalized by the value of the smallest price increment in EUR to reflect the trading standard unit 'value per traded lot'.

humans and the ML system can form a second-order mutual learning cycle: Human traders use their experiences and models to *frame* the ML environment by defining actions, context variables, and objectives that they consider relevant for trading. The ML system then acts and learns within this framed environment and *informs* traders about its experiences and inferred model. If done right, this can turn into a virtuous cycle: The better humans understand trading, the better they can frame the ML environment to improve its learnings. In turn, the better the ML becomes, the better the system can inform human traders to improve their understanding of trading. If done wrong, however, this may also turn into a vicious cycle, inhibiting individual and mutual organizational learning processes. While the humans and the ML system follow these learning cycles, we can observe the following differences in their (mutual) learning behavior.

(1.) Human Trading. For decades, AllGI relied purely on human learning to guide its trading endeavors. Each of AllGI's human traders develops her/his own unique propensities to act following the explained learning cycle (Figure 6a). Due to the discussed limits of human learning, human trading is far from being perfect: A trader's experience is limited to her/his unique course of trading; that is, no one trader can sample all trading algorithms across all imaginable market states but can only draw from her/his past trading choices, working time, and market developments during her/his career. Moreover, the traders can only perceive these experiences bounded by the context information they can gather. To comprehend the current market state, AllGI's traders primarily rely on up-to-date data from financial data providers (e.g., Bloomberg) and market commentary and reports from Brokers that they mainly consume through visual analysis (e.g., charts showing insights about market developments). To enrich these insights, the traders also use information from various available media (e.g., financial news portals, Twitter) to set the observed trends into a broader context of potential impact factors (e.g., tweets with political relevance). Yet, they can only inform themselves within a limited time span as they must rapidly react to the ever-changing market to not miss any opportunities. To this end, a single trader's experience is far from sketching the full picture of current market states. Even though traders can view their historic performance in reports, the hundreds of executions performed every day paired with the extensive number of factors affecting market developments complicate human traders to generalize accurate heuristics to form reliable trading strategies.

(2.) Autonomous ML-based Trading. In addition to its human trading, AllGI built an ML system for its trading executions. The ML system is implemented as a reinforcement learning agent that learns and trades autonomously without any active human involvement, mimicking the human trading process. Following reinforcement learning modalities, the ML system collects its own purely data-based trading experience by interacting with its environment (i.e., choosing trading algorithms within current market states). The ML system then uses the data-based experiences to infer heuristics about trading algorithms' success per market state which it uses to guide future actions (see also Figure 6b). Thereby, we refer to the *sample density* as the number of samples used by the ML system to learn about a specific state-action pair. In particular, AllGI's learning agent is implemented as a Q-learner that continually learns a Q-table to compute heuristics about pairs of trading algorithms and market states (Watkins & Dayan, 1992). Using Q-learning, trading is framed as a multiarmed bandit learning problem (Auer et al., 2002). Although the ML system acts and learns autonomously, its learning depends on human learning: To enlarge its own gathered experience, the ML system includes data of the humans' trades. AllGI further relies on its human traders' expertise to frame the ML environment, including the executable actions, reward function, and representation of market states. The traders preselected 22 trading algorithms as the ML system's executable actions based on an analysis of algorithm providers' reliability, historical execution performance, completion mechanics, personal preferences, and experiences. By doing so, the traders aim to exclude trading algorithms that appear generally unlikely to be successful and also disable an algorithm's use for specific market states when they agree on an algorithm being unlikely successful in these. With this framing of actions, the traders aim to reduce the necessity for the ML system to explore algorithm-state pairs that are very likely to produce low trading performance. The traders also defined the above-explained trading performance as the ML system's reward function to align its optimization objective with their own. Lastly, the traders also frame how the ML system perceives the current market state. In its current version, the traders framed market states along four continuous variables which describe the market state during a short time period before an order arrives at the trading desk: *impact_measure*, *mean_spread*, *std_deviation,* and *trade_frequency*. The *impact_measure* indicates the impact of the order relative to the market liquidity. The higher its value, the stronger a trade generally impacts market development. *Mean_spread* is the average distance between the bid and offer price. The higher the distance, the higher the leeway for trading becomes since each level between bid and offer can be used for trading and quoting. The *std_deviation* is the standard deviation of bid and offer prices, where a higher deviation indicates greater volatility in quoted prices. *Trade_frequency* is the average number of trades per minute in the market, reflecting how frequently other market participants trade. As the impact of others' trades can be observed, a higher *trade_frequency* facilitates to comprehend and hide the potential impact of own trades. As perceived by human traders, a security is more difficult to trade with increasing value of the variables (for *trade_frequency,* the opposite applies). Besides these variables, the ML system cannot perceive any other information about the market and therefore only focuses on these market characteristics, disregarding any other insights. At the time of our study, the ML system learned from 5493 self-executed and a thousand human trades. AllGI uses the ML system productively to trade an average amount of 10.2 billion EUR in notional per month.

(3.) Bilateral Human-Machine Trading. In the initial ML-based trading setup (as described above), the human trader's learning outcomes already affect the ML system as human expertise is used to frame the ML. To allow human traders to also learn from the ML system and thus to enable bilateral learning, AllGI enabled its ML system to advise human traders: as soon as a human trader assigns a new order, the ML system applies its model to the human's order without executing the order. The results of the applied ML model are then visualized for the human trader in a pop-up window. Figure 7a shows an exemplary visualization of the ML system's advice. Executable trading algorithms (anonymized as 'trade_algo_x' for this publication) are ordered along the X-axis based on their performance and a bubble's size represents the respective sample density. The Y-axis represents the algorithms' average expected performance with regards to the current market state. Next to the chart, the market state is described in a simplified form[10] to share the ML system's perception of the current market state. Human traders can also reopen all past advice from a centralized intranet website to revisit the ML system's advice without having to face a currently assigned order. To balance the trading allocation between the human and ML system, AllGI now lets the ML system randomly skip one-third of the trades that it would have usually traded. This allows us to investigate bilateral learning between machine and human on comparable datasets.

[10] The ML system's four continuous market variables are converted to simplified nominal values (i.e., low, low-med, med-high, high).

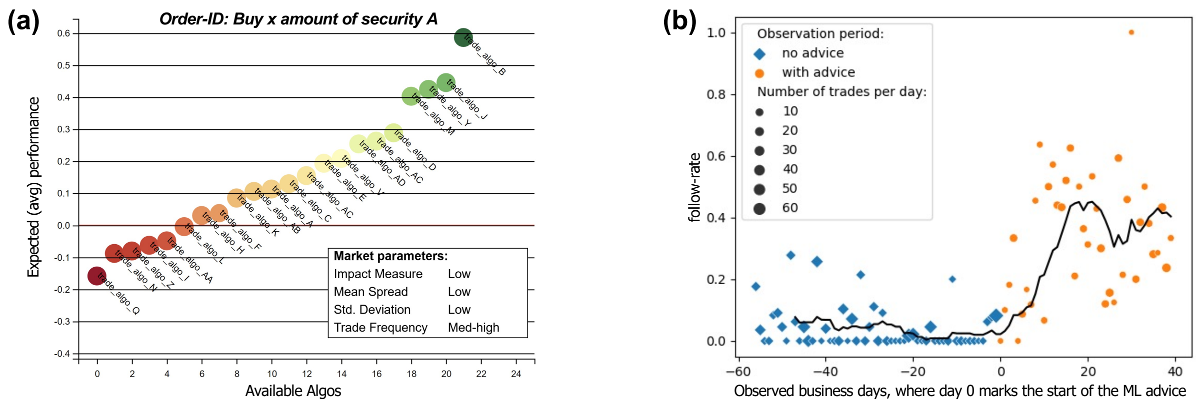

Figure 7: (a) Exemplary Trading Advice of the ML System and
(b) Daily Follow-Rate Before and After Enabled ML System's Advice

5.4.2 The Impact of Human and Machine Learning on Trading Performance

To understand how bilateral learning between AllGI's human traders and ML system affects trading performance, we now explore the behavior of AllGI's human traders when receiving trading advice from the ML system as additional guidance for each trade they execute.

Advice Consideration. To understand whether AllGI's human traders actually considered the ML system's trading advice, we compare the human trading decisions before and after enabled advice. We examine the *follow-rate* (i.e., percentage of trades where human decisions equal ML system's advice) to analyze whether the humans' and ML system's behavior became more alike after enabling the advice (we let the ML system give us post-hoc advice for all human trades prior to its enablement). We can cluster the data on the follow-rate into two categories: trading days prior to (56 days) and after (40 days) the enabled ML system's advice. Figure 7b shows the daily human follow-rate and a 10-business day moving average of the follow-rate to highlight the observable trend. The higher the follow-rate, the more the human traders took the ML system's advice on a given day. As implied by the point cloud in the upper right, human trading behavior tends to stronger align with the ML system's behavior after enabling its advice. On average, human trading equaled the ML system's advice in 5.3% of trades prior to and in 34.0% after the enabled advice.[11]

Performance Impact. To understand whether the ML advice helps AllGIs' human traders, we now nuance our analyses of the human follow rate. We calculated the *advice distance* for each advice, which reflects a normed performance difference between a taken decision and received advice in relation to the best and worst decision.[12] We use the advice distance as an independent input variable, which we compare against trading performance. A human trader may hold broader knowledge or intuition about market developments which the ML system cannot quantify with its narrow market view. If a human trader believes another trading choice to be more favorable, the trader will likely neglect the ML system's advice. If traders do this well, human trading decisions should not positively correlate with the advice distance. Figure 8a

[11] We also computed Kendall's tau-b correlation coefficient (Daniel, 1990; Kendall, 1945) between the provision of ML advice and the daily average percentage of humans following the ML advice for 96 days. We found a strong, positive association between providing ML advice to the humans and the humans following the advice, which is statistically significant with τ_b = .653 and p = .0005.

[12] We defined the advice distance as $d_{Advice} = 1 - (score_{BestDecision} - score_{MadeDecision})/(score_{BestDecision} - score_{WorstDecision})$. Orders for which the selected algorithm was beyond the ML system's set of algorithms were not considered for this analysis.

illustrates performance and related advice distances of the human trades. To highlight contained trends, we clustered the trades into five groups along the advice distance: in addition to trades with followed advice (i.e., advice distance = 0), we split the no-follows (i.e., advice distance > 0) into quantiles along the advice distance. Figure 8b shows that, with increasing advice distance, the boxplots of each group shift downward toward lower performance, as do their medians (i.e., 5.3, 2.4, .01, .0, .0).[13] While the effect appears small, it should not be neglected as such small improvements in trading already create large profit gains. In sum, the more the human traders follow the ML system's advice, the more effective their trading appears to become.

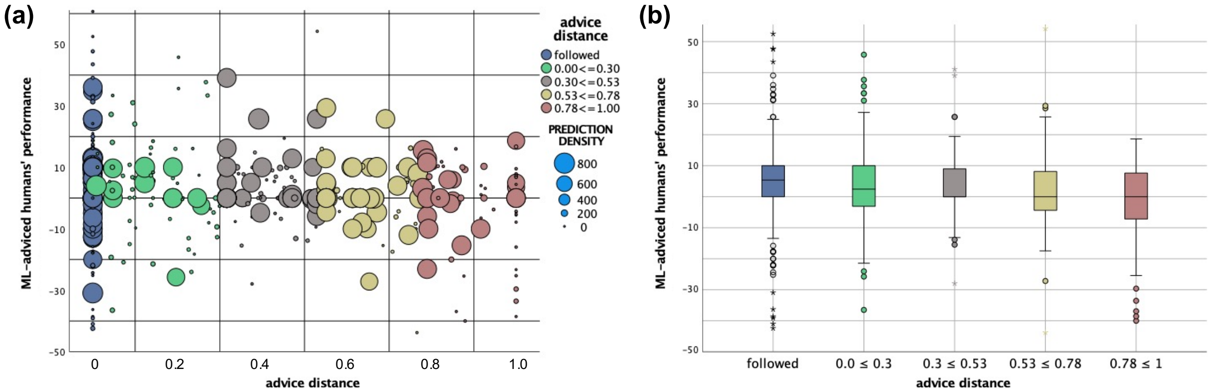

Figure 8: (a) Human Trades (N = 561) Plotted Along Performance and Advice Distance
(b) Boxplots of Human Trades Clustered into Followed and No-Follow Quantiles

Yet, although following the ML system's advice tends to improve human performance, the ML system does not outperform the human traders; that is, on average, human traders and the ML system perform equally well after all.[14] However, if we regard the joint performance of the humans and ML system before (during the first 56 business days) and after (during the subsequent 40 business days) the enabled advice, we can observe that AllGI's overall trading performance increased significantly.[15] The enabled bilateral human-machine learning appears to have improved AllGI's trading performance.

5.4.3 *Unveiling the Mutual Dynamics within Bilateral Human-Machine Learning*

At first glance, our observations appear rather paradoxical: AllGI's ML system does not outperform the human traders but the human traders improve their performance if they follow the ML system's advice more closely. To better understand the human-machine dynamics that underlie this conundrum, we now take a closer look into both actors' trading behaviors along different market scenarios.

---

[13] We also computed a Spearman's rank-order correlation (Daniel, 1990; Spearman, 1987) between advice distance and performance of an order. We found a weak negative correlation between advice distance and performance (i.e., performance decreases as advice distance increases), with $rs(559) = -.176$ and $p = .0005$.

[14] We ran a Mann-Whitney U test (Daniel, 1990; Mann & Whitney, 1947) to assess differences in trading performance between ML-assisted humans and the ML system while the ML system's advice was enabled (N = 1871 trades). Distributions of the performance for ML-assisted humans and the ML system were not similar, as assessed by visual inspection. The performance for the ML-assisted humans (mean rank = 920.42) and the ML system (mean rank = 944.70) were not statistically significantly different, with $U = 414028.5$, $z = -.962$, and $p = .336$.

[15] We ran a Mann-Whitney U test (Daniel, 1990; Mann & Whitney, 1947) to assess differences in trading performance before and during enabled ML system's advice is statistically significant (N = 3797 trades). We found that performance distributions before and during enabled ML system's advice were not similar. The performance during enabled ML system's advice (mean rank = 1947.94) was statistically significantly higher than before (mean rank = 1851.61), with U = 1893097, $z = 2.714$, and $p = .007$.

The violin plots in Figure 9 show how human performance is distributed along four key market dimensions (i.e., the variables also used by the ML system), each with four different levels ranging from low to high. Each 'violin' is split into two colors: The orange area shows the performance distribution along the Y-axis of the humans when following and the blue area when not following the ML system's advice. At the same time, the bulge size of the colored areas indicates the underlying amount of data. When comparing the violins within each dimension, we can observe two tendencies: First, as the level increases from low to high, the amount of follows relative to no-follows decreases (exception: for *trade_frequency,* we can observe the opposite trend). Remember that with increasing levels of the variables, trading becomes more difficult for the traders (for *trade_frequency* the opposite applies). Thus, the more difficult trading becomes, the less the human traders appear to follow the ML system's advice. Second, for all levels, the orange distributions' peaks are skewed stronger towards the top than the blue ones, indicating a better overall performance when humans follow the advice. This observation is consistent with our analysis in the previous section: when following the ML system's advice, AllGI's human traders tend to improve their performance.

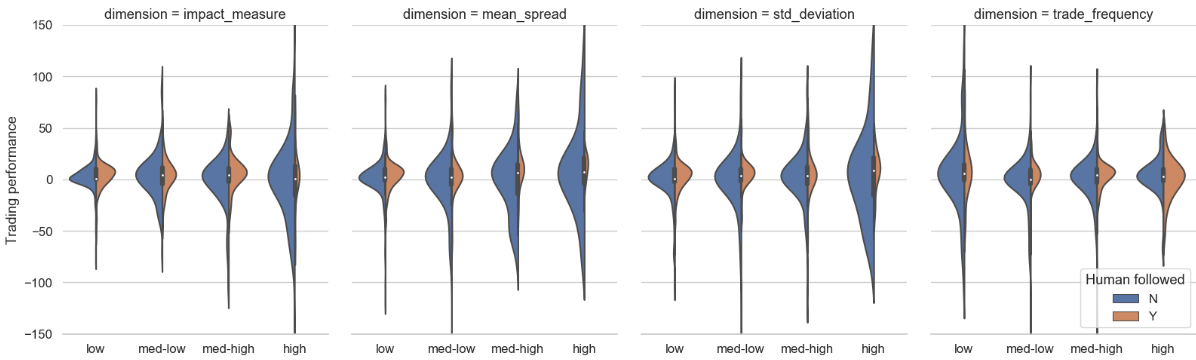

Figure 9: Performance Distributions Along Different Market Scenarios

To better understand why the human traders do or do not follow the ML system's advice in certain scenarios, we talked to the human traders on how they utilize the ML system's advice. The traders agreed that, when receiving such advice, they essentially try to reflect on their own and the ML system's gathered amount of experience with the faced scenario. Note that simple scenarios also reflect scenarios that are traded very frequently at AllGI. Both the human traders and the ML system have therefore already gathered related experiences extensively, and, due to the great sample combined with the relatively limitedly faced difficulty, already developed their own strong beliefs about how to trade such scenarios. While they mostly favor their own developed trading strategies that appear most reliable in the face of their past experiences, AllGI incentivizes its human traders to keep improving their trading through exploration of new trading strategies, as their bonus is coupled with increased trading performance. As the ML system's advice represents approaches that were successful in the past and condenses its and multiple traders' experiences, it grants human traders the possibility to occasionally explore different but equally reliable alternatives, reducing the risk of costly failures that exploration of new approaches usually comes with. By doing so, the ML system helps to break open solidified (potentially suboptimal) propensities, helping humans to reevaluate and nuance their own developed strategies. This is in contrast to more difficult, rarer scenarios, where the human traders and the ML system have so far only gained limited experience to develop robust trading strategies. In such scenarios, the human traders still explore extensively to uncover more reliable strategies. Here, deliberately reflecting on experiences is especially relevant to not only identify promising opportunities but to also bypass ominous actions as flawed or needlessly repeated explorations can accumulate several million euros in trading costs. Thus, to better guide their exploration, human traders utilize the ML system's advice to not simply follow the so-far most promising approach but primarily to reflect on organization-wide experiences. This allows them to either further explore seemingly promising approaches or identify unexplored 'blind spots' that might yield further beneficial directions. Consequently, the human traders tend to follow the ML system's advice less frequently in more difficult trading scenarios.[16]

While this shows how ML affects human trading behavior, we now explore how human trading affects the ML system's learning over time (remember: the ML system also learns from human trades). The sankey diagrams in Figure 10 outline the ML system's behavioral changes in relation to the humans' trading behavior: Both diagrams consist of three segments, respectively showing the aggregated decisions regarding individual trading algorithms by the human traders (left), the accordingly received ML system's advice (middle), and the changed advice that the ML system provides for these human trades at the end of the observation period when this human data is used for retraining the ML system (right). Each node (represented by a letter and a color box) groups trades where a specific trading algorithm (e.g., *'D'*) was used or advised. The links between the nodes connect single trades across all three segments. The larger a node, the more frequently the respective trading algorithm was used or advised. The thicker a link, the more frequently respective connections appeared. Consequently, paths between the left and middle segment represent whether a human followed an advice or not, while paths between the middle and right segment represent changes in the ML system's advice when being retrained with all human trading choices. While the left diagram (a) bases on all observed trades, the right diagram (b) shows an excerpt covering only trades in simple scenarios (i.e., frequent trades for which the ML system learned from a high sample density).

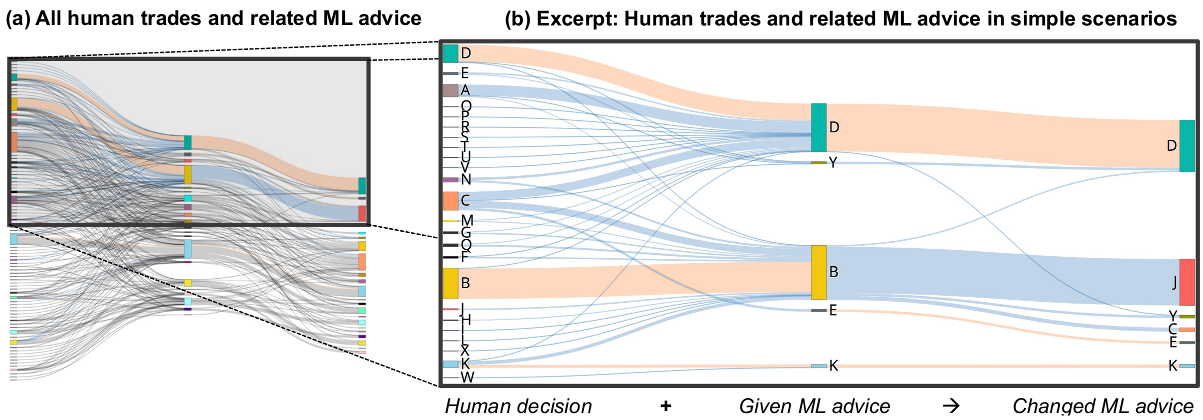

Figure 10: (a) All Human Trades (Left) Linked to Given (Middle) and Eventually Changed (Right) ML System's Advice; (b) Excerpt of Trades in Simple Scenarios

In Figure 10, the higher number of nodes in the human segments demonstrates that the humans used a larger variation of trading algorithms. When focusing on trades of simple scenarios (Figure 10b), we can clearly observe that human trading affected the ML system's trading behavior in several ways: The ML system revised its strongest strategy from using trading algorithm B to J and diversified it with the use of the additional algorithms Y and C in certain scenarios. Although in 65% of trades the human traders used a variety of other algorithms (blue

---

[16] We also computed Kendall's tau-b correlation coefficient (Daniel, 1990; Kendall, 1945) between following the ML system's advice (Yes = 1/No = 0) and the sample density of the advice (N = 766). We found a positive association between humans following the ML system's advice and a higher sample density, which is statistically significant with τ_b = .170 and p = .0005.

links) instead of following the advice D, the ML system reinforced and extended its use of the superior algorithm D. While the human traders increasingly explored algorithms A and C in various scenarios, the ML system figured to keep both in most cases. In complex scenarios (the nodes with grey connections in Figure 10a), comparable patterns can be observed that, due to an even greater diversity in human trading and less experience of the ML system, yielded more changes in the ML system's advice. To this end, the vibrant human trading helps the ML system to enrich its experiences which stimulates a broader assessment, revision, and extension of its inferred trading strategies.

5.5 Discussion

With the increasing use of ML systems in organizational routines alongside human counterparts, enabling effective human-machine collaboration becomes ever more important. Only recently, scholars started to stress the great importance of managing the bilateral relationship between humans and ML systems (e.g., Rai et al., 2019; Schuetz & Venkatesh, 2020) and increasingly acknowledge the great potential of organizational learning research to analyze this relationship (e.g., Lyytinen et al., 2021; Ransbotham et al., 2020; Seidel et al., 2019). Yet, despite decades of research on organizational learning, scholars have largely assumed the learners to be purely human (e.g., Argote et al., 2021; Argote & Miron-Spektor, 2011). So far, only a handful of scholars have studied potential impacts of ML on organizational learning but mainly remained on a theoretical level and call for further research (i.e., Afiouni-Monla, 2019; Balasubramanian et al., 2022; Lyytinen et al., 2021; Ransbotham et al., 2020; Seidel et al., 2019; Sturm, Gerlach, et al., 2021). With our study, we aim to help answer these calls by drawing on a case in a real-world organizational context. We provide empirical insights to enrich ongoing discussions, hoping to inspire further research endeavors and help organizations design effective human-AI collaborations.

Our study contributes to theory in multiple ways. First, given its impact on organizational performance, our study emphasizes that an organizational learning perspective should not be neglected when managing the emerging bilateral human-AI dynamics. With our case, we demonstrate how scholars and organizations can adopt this perspective to empirically identify and explain actual (interrelated) behaviors of humans and ML systems within organizational contexts. Second, hoping to inform such studies, we condensed key idiosyncrasies of human learning and ML from existent literature to help recognize potentials for change when human learning is being replaced or augmented with ML. Moreover, drawing on these theoretical idiosyncrasies and insights from our case, we proposed a conceptualization of the fundamental learning processes and linkage between humans and ML systems. As shown in our case study, both can be used to theoretically ground agents' characteristics and relations to guide empirical analyses of reciprocal human-AI dynamics and explain their (unintended) consequences. Third, as we based our analyses on a novel method of human-machine pattern recognition ourselves, our study demonstrates how human-AI collaboration can also benefit research endeavors. Scholars interested in leveraging this potential can rely on our study to stimulate comparable research designs. Our study further illustrates how especially digital trace data analyses can thereby act as a powerful and context-rich tool to help unravel complex behavioral dynamics, allowing to (1) leverage machines' high information processing capabilities for recognizing fruitful patterns in extensively tracked activities and (2) humans' broader contextualization capabilities for finding explanatory patterns through inquiries with involved actors and contexts. Our case exemplifies that one must be careful with isolated analyses when aiming to unravel ML's complex consequences, urging scholars to consider varying temporal or spatial foci when studying human-AI relationships.

Fourth, our study adds empirical insights to the ongoing automation-augmentation discussion on whether humans should be taken 'out of the loop' as soon as ML systems can reliably replace them in their routines (e.g., Brynjolfsson & Mitchell, 2017; Raisch & Krakowski, 2021). Although AllGI's ML system enables successful autonomous learning, our observations suggest that AllGI can benefit from keeping its human traders 'in the loop' as their broader contextualization adds value to the whole learning system. Due to its reliance on a preselected subset of trading algorithms and market variables, the ML system can explore and infer trading strategies only within its narrow frame. Despite its less bounded information processing, an increased reliance on the ML system's trading would therefore come with the risk of restricting AllGI's trading strategy innovations to the limits of the ML system's framed view. With their holistic market view and less rational behavior, the inclusion of humans helps AllGI to actively counteract this risk as they help to look 'outside the box'; that is, while the ML system grants AllGI the ability to learn reliable strategies within its boundaries, AllGI requires its human traders to learn about promising algorithms and pivotal market conditions beyond the system's frame. Only if AllGI's human traders keep actively learning through their own trading experiences, they can translate the complex, ever-changing trading environment to a substantiated, fruitful frame in which the ML system can unfold its preeminent learning capabilities.

Lastly, our study further contributes to the emergent discussion on whether ML amplifies (e.g., Balasubramanian et al., 2022) or alleviates (e.g., Sturm, Gerlach, et al., 2021) organizations' learning myopia (i.e., the tendency to favor exploitation over exploration; Levinthal & March, 1993). In the case of AllGI, we can observe that the human traders take a more explorative role while the ML system tends to act rather exploitatively. For AllGI, increasingly shifting trades to the ML system thus appears to come with an increased risk of stagnation when ML outcomes remain isolated. However, when enabling humans to also learn from the ML system (i.e., enabling bilateral learning), we can observe an opposite effect; that is, human exploration benefits from the ML system's increased exploitative behavior as its shared learnings allow humans to better reflect on past organization-wide experiences and help to rationally uncover blind spots and promising strategies. Depending on the maturity of the developed beliefs, ML thereby either helps the humans to break open solidified propensities or better guide ongoing explorations of so-far uncharted areas. In contrast, the human traders' boundedly rational trading behavior adds larger variations to the ML system's experience which would have been neglected within its narrowly focused formal exploration, helping to overcome potential convergence towards suboptimal strategies. To this end, our findings point to a potential virtuous cycle between human learning and ML, in which humans improve ML through diversifying experiences for the ML system's exploitation and ML benefits human learning through informing human exploration. These insights demonstrate that well-coordinated bilateral human-machine learning can act as an effective mechanism for organizations to counteract myopia. To make use of this potential, future research could explore organizational designs on how to unite both learners' idiosyncrasies to mutually enhance and effectively balance explorative and exploitative behavior and focus on more detailed analyses of how different contexts (e.g., varying amount of experience or faced complexity) affect learning myopia in human-machine collaborations. Especially with myopia's broad context dependency, future research can help uncover the extent to which observed effects are unique to the trading context's specific decisions, objectives, and incentives, and how they can be translated to divergent contexts, such as more subjective, more ambiguous, or less competitive domains (e.g., the health care or public sector). We also encourage scholars to explore conditions and mechanisms that sustain the identified dynamics in the long term or impede them, in the worst case even turning bilateral learning into a vicious cycle.

Our insights are equally relevant for practitioners. Organizations can use our results to inspire and design future AI initiatives beyond ML-based automation of routines. To leverage ML's full potential, our study demonstrates that practitioners are well-advised to consider potential human-AI collaboration designs already at an early stage of their ML development. Our observations should encourage organizations to rethink whether integrating ML within their routines gave birth to new knowledge silos, how they can make this ML-based expertise available to human experts, and how they can maintain human-AI knowledge transfer in both directions to enable synergy effects. In its essence, this is not only a matter of organizational design, but reflects a new managerial issue that requires organizations to deliberately reflect on both actors' idiosyncrasies to effectively coordinate their emerging dynamics.

Of course, our study is subject to limitations. Due to our study's exploratory nature, future research should validate and contextualize our findings under different boundary conditions across heterogeneous contexts and domains. As we relied on a single case, our observations must be interpreted within the limits of its peculiarities. It is not clear that similar dynamics emerge in comparable routines, especially if other ML types, human-AI collaboration setups, or observation periods are used. Although we aimed to maximize the breadth of our observations by drawing on a rich and diverse set of data, there are certainly many aspects that we could not observe but could still influence human-machine dynamics since they were not trackable in digital traces or not obvious to the traders. Here, we invite future studies to help analyze further aspects, such as motives or activities, that remained hidden to us by relying on different contexts, methods, or foci.

5.6 Acknowledgements

This study has been funded by the German Federal and Hessen State Ministry for Education and Research within their support of the National Research Center for Applied Cybersecurity ATHENE.

6 Paper D: Artificial Intelligence and Organization-level Performance (Focus: Organizational Learning)

Title

Coordinating Human and Machine Learning for Effective Organizational Learning

Authors

Sturm, Timo; Gerlach, Jin P.; Pumplun, Luisa; Mesbah, Neda; Peters, Felix; Tauchert, Christoph; Nan, Ning; and Buxmann, Peter

Publication Outlet

Management Information Systems Quarterly (MIS Quarterly), Vol. 45 No. 3, Sep. 2021, pp. 1581–1602. DOI: 10.25300/MISQ/2021/16543

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

Please note:

Due to copyright restrictions, only the first paragraph of the discussion section and the table summarizing the main findings, propositions, and implications of our study have been included. Please refer to the article published in the MIS Quarterly journal to access the study in its original and complete form.

6.1 Discussion

In organizations where ML is increasingly taking over task responsibilities, coordinating humans and ML systems has become a challenging management issue. Our research highlights that the outcomes of organizational learning must not be neglected when coordinating the work of humans and ML systems, given the profound consequences of an organization's long-term knowledge. Table 4 summarizes our study's main findings and propositions (P1, P2, and P3) as well as their implications.

Table 4: 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 Contributions and Conclusion

To achieve great performance, organizations must make sense of the reality that surrounds them so that they can effectively act in and adapt to their environment (e.g., Argote & Miron-Spektor, 2011; March, 1991, 2010). To date, organizations have exclusively relied on their human members as the only actors capable of learning (e.g., Argote et al., 2021; Argote & Miron-Spektor, 2011; Levinthal & Rerup, 2021), and the limits of human cognition have thus complicated organizations' learning since the very beginning (e.g., Levinthal & March, 1993; March, 2010; Simon, 1972). Hoping to overcome human limitations (e.g., in their attention to and comprehension of aspects of reality; March, 2010; Simon, 1972), today's organizations leverage ML systems as powerful new learners alongside humans (e.g., Balasubramanian et al., 2022; Berente et al., 2021; Ransbotham et al., 2020; Seidel et al., 2019). However, ML systems do not necessarily represent a better but rather a *different* form of learner that provides its own advantageous and disadvantageous idiosyncrasies (e.g., Argote et al., 2021; Balasubramanian et al., 2022; Berente et al., 2021; Ransbotham et al., 2020; Sturm, Gerlach, et al., 2021). These differences may provide room for synergies when humans' and ML systems' abilities are combined (e.g., Lyytinen et al., 2021; Murray et al., 2021; Ransbotham et al., 2020; Schuetz & Venkatesh, 2020; Tremblay et al., 2021). However, research remains broadly silent on how such synergies can be enabled and on the consequences that may arise from joint learning efforts between humans and ML systems.

In this dissertation, I aim to help advance our understanding of the interplay between humans and ML systems by uncovering their virtuous and vicious learning dynamics and how organizations can effectively manage these dynamics to make deliberate use of them. To do this, I have explored how emergent dynamics and coordination designs affect performance on three levels of analysis, namely, the individual, group, and organization levels. Based on the four included studies, I offer the following contributions to research and practice.

7.1 Theoretical Contributions

Addressing current calls for research on human–AI collaboration (e.g., Argote et al., 2021; Baum & Haveman, 2020; Benbya et al., 2021; Berente et al., 2021; Murray et al., 2021; Rai et al., 2019; Schuetz & Venkatesh, 2020), this dissertation provides several theoretical contributions to help better understand and control the virtuous and vicious dynamics of ML systems in their interplay with humans in organizations.

First, the four studies explore performance impacts on the three levels of individuals, groups, and organizations. Overall, the studies demonstrate that positive performance impacts can be achieved on all three levels when combining humans and ML systems in organizations. As the studies also show, these positive findings require organizations to meet ML-specific conditions and to craft effective coordination designs. The studies further point to the adverse effects of ML systems on the overall performance of humans and organizations that can otherwise arise. In what follows, these essential insights are further nuanced along the three levels of analysis.

On the *individual level*, research has already acknowledged the potential value of delegating tasks to ML systems and of the interaction between humans and ML systems (e.g., Brynjolfsson & Mitchell, 2017; Rai et al., 2019; Raisch & Krakowski, 2021; Schuetz & Venkatesh, 2020). Yet current research offers only limited insights into the relevant factors that influence how tasks should be assigned between a human and/or an ML system and should therefore be considered by organizations to enable effective allocation and linkage of tasks in ML contexts (e.g., Brynjolfsson & Mitchell, 2017; Fügener, Grahl, Gupta, & Keter, 2021; Raisch & Krakowski, 2021; Schuetz & Venkatesh, 2020). To help extend our understanding of such factors, the studies included in this dissertation offer further insights regarding how organizations' task allocation decisions between a human and an ML system can be designed effectively. Two of the studies contribute to advancing our understanding of the conditions and design factors for (1) task delegation (paper A) and (2) forming collaborative links (paper B) between a human and an ML system. Paper A conceptualizes how organizations can uncover scenarios in which the idiosyncratic advantages of ML systems can unfold. The study identifies key factors and procedural artifacts that affect the identification and formation of contexts in which the use of an ML system instead of a human can benefit organizational performance. To build potential complementarities based on the delegated tasks, paper B captures manageable levers of interaction between a human and an ML system and highlights conditions under which such interaction can positively complement individual performance.

Regarding the *group level*, research on the impact of ML systems on group performance remains in its infancy, leaving it largely unclear how groups that include ML systems should be coordinated and what consequences may arise from different coordination strategies (e.g., Fügener, Grahl, Gupta, & Ketter, 2021; Grønsund & Aanestad, 2020). To help inform the formation of effective group coordination strategies, paper C demonstrates how virtuous and vicious dynamics between humans and an ML system can emerge and impact their group performance. The findings show that the bilateral interaction between multiple humans and an ML system can connect and promote their respective idiosyncrasies, yielding synergetic roles and performance impacts (e.g., an ML system that tends to exploit can be used to stimulate human exploration). The study reveals the performance gains that can be generated exclusively from the knowledge exchange between humans and an ML system.

On the *organization level*, performance impacts are difficult to observe, which strongly limits current research to pure theoretical analyses (one exception is Balasubramanian et al., 2022), rendering our understanding of the organization-wide dynamics of human–machine collaborations scarce (e.g., Baum & Haveman, 2020; Berente et al., 2021; K. Leavitt et al., 2021; Lindebaum et al., 2020; Murray et al., 2021; Raisch & Krakowski, 2021). Paper D offers rare observations of such dynamics and outlines ML systems' potential to disrupt organization-wide routines and norms by contributing their own knowledge to the organization's stock of knowledge. The study shows that a wide-ranging use of ML systems can, indeed, improve an organization's overall performance. However, this requires effective coordination of the contributions of *all* humans and ML systems as well as their interplay while considering the environmental conditions. Otherwise, ML systems can harm human and organizational knowledge creation, which can have wide-ranging consequences for an organization's overall evolution and survival.

The findings of the four studies thus provide ground for theorizing about the conditions, complementarities, roles, dynamics, and consequences that should be considered when coordinating humans and ML systems on the various organizational levels, which I hope will stimulate further studies on human–machine collaborations in organizations.

Second, to better understand the complementarities and incompatibilities between humans and ML systems, the studies unpack their respective idiosyncrasies. Conceptualizations of problems that are suited for ML systems (paper A), conditions for complementarities between humans and ML systems (paper B), bilateral human–machine learning cycles (paper C), and human (re)configuration of ML systems (paper D) are proposed to capture such idiosyncrasies. The strengths and weaknesses of ML systems are thereby empirically confirmed and nuanced. For instance, the studies empirically demonstrate how ML systems' less-bounded rationality and what is known as the "frame problem" can facilitate and limit the informativeness of the derived learning. The studies also show, for example, that the provision of comprehensive and accurate data is one of the greatest organizational challenges, rendering imperfect data the norm and an important condition to consider when managing the use of ML systems and their interaction with humans. To inspire future research endeavors, these conceptualizations can be used to focus studies on aspects that are unique to ML systems relative to traditional (non-ML) ISs and humans. This may help in reflecting on whether and to what extent ML systems make it necessary to revise existing theories or to theorize in new, unorthodox ways.

Third, although ML systems exhibit autonomy, the results of these studies stress that humans must be involved in the ML systems' learning in the long term, even in the presence of highly effective ML systems. After all, it is humans who identify appropriate problems, select and prepare data to capture the problems, select and parametrize algorithms, and evaluate the created ML systems. To effectively perform these activities, humans require knowledge that enables them to assess the relevance of problems, describe problems in terms of relevant variables, and evaluate the quality of proposed problem solutions. The four studies underline that, in order to do so, humans need to remain "in the loop" and continue their learning in order to develop an up-to-date understanding of how to identify and solve relevant problems. The better humans understand a problem domain, the more likely they will be able to identify purposeful problems for ML systems, describe their most relevant aspects along with data, and optimize ML systems toward applicable solutions. Otherwise, ML systems may stop evolving, rendering them obsolete, and increasingly drift far away from reality, which can be detrimental to human and organizational performance due to the potential stagnation in learning and the promotion of flawed beliefs.

Fourth, the studies also provide insights into the potential role of ML systems and the changing role of humans in organizations. The studies demonstrate empirically that ML systems can add value beyond pure automation if their derived insights are shared with humans. Papers C and D both illustrate that positive synergies between humans and ML systems can be achieved by creating reinforcing cycles of mutual improvement through effective coordination. The papers reveal two worthwhile role assignments within these cycles: Paper C uncovers that, if ML systems exploit while humans explore, ML systems can help humans better guide their exploration (e.g., through highlighting blind spots or promising/unpromising directions) and break apart solidified propensities to act, while the vibrant human learning can help ML systems to diversify their exploitation. Paper D highlights that, in light of humans' natural tendency to exploit (i.e., their learning myopia), ML systems can also be used to explore in order to liberate humans from the need to explore without sacrificing organizational performance. Both role assignments require organizations to view ML systems not as substitutes or dull support, but rather as active collaborators with humans to allow for joint knowledge creation. The studies illustrate that, if coordinated effectively, such collaboration can surpass the outcomes of either purely human or purely ML system scenarios on the different levels of analyses. Despite this positive observation, the studies also foreshadow vicious cycles that may arise from flawed coordination designs and role assignments that create reciprocal impediments (e.g., if ML systems only imitate and strongly reinforce historic human behavior, they can impede the human exploration that is also required for worthwhile revisions of ML over time, potentially leading to a stagnation of the mutual learning cycle; Sturm, Koppe, et al., 2021).

Fifth, by unpacking the bilateral relationship between humans and ML systems, the studies further show that the value created through their interactions arises from sharing their *different* approaches to solving the same problem. For instance, in paper C, it was precisely the bilateral sharing of different approaches when collaborating that helped the humans better guide their exploration and the ML system better reevaluate its exploitation. Comparable dependencies and dynamics also arose on the organization level in paper D. This reveals a fundamental dilemma that poses a new managerial issue: While value is created from the difference in the insights that are shared by humans and ML systems, the sharing of insights may add value in the short term but jeopardizes long-term value as humans and ML systems are likely to become more similar over time when gradually adopting each other's insights. Therefore, an important question remains for future research regarding how organizations can connect the learning of their humans and ML systems to exploit complementarities while preserving the uniqueness of their individual approaches. While the study in paper D provides initial evidence and coordination designs to address this issue, more research is required to validate, extend, and translate the findings to different real-world contexts.

7.2 Practical Contributions

The four studies also point to guiding principles that can help practitioners make effective use of ML systems. First, the study results can help practitioners select and form their future ML initiatives. Practitioners can use the findings to identify suitable problems and create conditions that can enable and facilitate the use of ML systems in a structured manner. While organizations will thereby be enabled to evaluate the feasibility of possible ML initiatives, the insights provided by this dissertation can also help protect them from mistakenly promoting initiatives that are ill-suited for ML systems and that may in fact inhibit otherwise achievable performance.

Second, promoting the perspective of organizations as complex systems of interacting humans and ML systems can also help practitioners optimize the integration of ML systems into their organizational processes. Rather than fully investing scarce resources in the development of highly optimized ML systems, practitioners should not neglect the design of bilateral relationships between humans and ML systems. Following Lyles and Fiol's (1985) notion that "organizational learning is not simply the sum of each member's learning" (p. 804), the studies demonstrate that the very interaction between humans and ML systems can itself create value. Shifting resources into enabling such effective interactions (e.g., by enabling the exchange of data and insights between the two actors) may therefore benefit both humans and ML systems, improving their overall collaboration and mutual performance.

Third, the value of the interaction between humans and ML systems shows that practitioners should carefully consider whether it is worthwhile to replace human experts with ML systems.

Despite the short-term benefits of ML-based automation of human tasks, the studies included in this dissertation stress that human expertise continues to play an important role in the effectiveness of ML systems' long-term reconfiguration. Practitioners would therefore be well-advised to allow some of their human experts to continue to work side by side with ML systems, even in the presence of highly advanced ML systems. Further advancement of ML systems can therefore require an organization to further invest in its human *domain* expertise in addition to technical expertise and infrastructure. For instance, this may include intensified training of human domain experts, further investments in R&D, and acquisition of external expertise through personnel turnovers within the domains that are already covered by ML systems.

Fourth, the findings also stress that although simply enabling interactions between humans and ML systems is an important primary requirement for collaborations, it may not be sufficient for exploiting their full potential. Such collaborations add value through the learning cycles that emerge between humans and ML systems, creating dynamics that generate either positive or negative reinforcements within organizations. In the best case, these collaborations can yield virtuous learning cycles between humans and ML systems that enable the mutual enhancement of both actors (e.g., through the ML systems informing humans about as-yet overlooked patterns and the humans using their improved knowledge to better reconfigure the ML systems). In the worst case, however, collaborations may also create vicious cycles that harm the performance and advancement of both actors (e.g., if incorrect beliefs, such as prejudices, are promoted by ML systems and are then continuously reinforced in the ML systems by human reconfigurations). The studies demonstrate that control of such dynamics is not only a technical challenge but primarily poses a managerial issue: the coordination of humans, ML systems, and their interplay. Here, the studies show that effective coordination involves placing humans and ML systems into roles that can promote their unique abilities (e.g., human imagination and ML systems' less-bounded rationality) to enable worthwhile synergies. Moreover, coordination also raises the challenge of facilitating the exchange of both actors' contributions while preserving their differences in order to enable synergies through complementarities. To achieve this, practitioners should not try to always have their ML systems completely imitate their human experts but should rather grant them some degree of freedom to allow for approaches that differ from human-made solutions. For instance, channeling different information to both actor types and varying their incentives for taking action may be fruitful mechanisms to enable and preserve the development of different but potentially complementary behaviors.

7.3 Concluding Remarks

ML systems are neither a universal panacea for understanding reality nor a curse that threatens the relevance of human cognition. Humans and ML systems make *different* kinds of intelligence available to organizations, with neither promising generally superior performance. It is precisely the unique differences between humans and ML systems that often seem to make them better complements than substitutes. The secret to unleashing the true potential of ML systems may therefore lie in effectively coordinating the differences between humans and ML systems and their bilateral relationship to produce virtuous cycles of mutual improvement. This dissertation is a first step toward developing theory and guidance for such worthwhile collaborations, but more work is needed to rethink collaboration theory in the era of AI.

References

Afiouni-Monla, R. (2019). Organizational Learning in the Rise of Machine Learning. *Proceedings of the 40th International Conference on Information Systems*.

Afrashteh, S., Davern, M., & Someh, I. A. (2020). Enhancing Fairness in Algorithmic Decision-Making Through Perspective Taking. *Proceedings of the 28th European Conference on Information Systems*.

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

Agrawal, A., Gans, J. S., & Goldfarb, A. (2017). What to Expect From Artificial Intelligence. *MIT Sloan Management Review*, *58*(3). http://mitsmr.com/2jZdf1Y

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. http://www.jstor.org/stable/3250961

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

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine Bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed on Oct 30, 2022)

Argote, L., Lee, S., & Park, J. (2021). Organizational Learning Processes and Outcomes: Major Findings and Future Research Directions. *Management Science*, *67*(9), 5399–5429. https://doi.org/10.1287/mnsc.2020.3693

Argote, L., & Miron-Spektor, E. (2011). Organizational learning: From experience to knowledge. *Organization Science*, *22*(5), 1123–1137. https://doi.org/10.1287/orsc.1100.0621

Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, *47*(2), 235–256. https://doi.org/10.1023/A:1013689704352

Balasubramanian, N., Ye, Y., & Xu, M. (2022). Substituting Human Decision-Making with Machine Learning: Implications for Organizational Learning. *Academy of Management Review*, *47*(3), 448–465. https://doi.org/10.5465/amr.2019.0470

Basadur, M. S., Graen, G. B., & Green, S. G. (1982). Training in creative problem solving in an industrial research organization. *Organizational Behavior and Human Performance*, *30*(1), 41–70. https://doi.org/10.1016/0030-5073(82)90233-1

Baum, J. A. C., & Haveman, H. A. (2020). Editors' comments: The future of organizational theory. *Academy of Management Review*, *45*(2), 268–272. https://doi.org/10.5465/amr.2020.0030

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/#7123b2606016 (accessed on Oct 30, 2022)

Benbya, H., Nan, N., Tanriverdi, H., & Yoo, Y. (2020). Complexity and information systems research in the emerging digital world. *MIS Quarterly*, *44*(1), 1–17. https://doi.org/10.25300/MISQ/2020/13304

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

Benta, C., Wilberg, J., Hollauer, C., & Omer, M. (2017). Process model for data-driven business model generation. *Proceedings of the 21st International Conference on Engineering Design*, 347–356.

Berente, N., Gu, B., Recker, J., & Santanam, R. (2021). Special Issue Editor's Comments: Managing Artificial Intelligence. *MIS Quarterly*, *45*(3), 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Berthelot, D., Oliver, A., Carlini, N., Goodfellow, I., Raffel, C., & Papernot, N. (2019). MixMatch: A Holistic Approach to Semi-Supervised Learning. *Proceedings of the 33rd Conference on Neural Information Processing Systems*, 5049–5059.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer Science and Business Media.

Björk, J. (2012). Knowledge domain spanners in ideation. *Creativity and Innovation Management*, *21*(1), 17–27. https://doi.org/10.1111/j.1467-8691.2012.00627.x

Bremser, C. (2018). Starting points for big data adoption. *Proceedings of the 26th European Conference on Information Systems*.

Brynjolfsson, E., & Mcafee, A. (2016). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. Norton & Company.

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

Burks, A. (1946). Peirce's Theory of Signs. *Philosophy of Science*, *13*(4), 301–306. https://doi.org/10.1086/286904

Chai, Y., & Li, W. (2019). Towards Deep Learning Interpretability: A Topic Modeling Approach. *Proceedings of the 40th International Conference on Information Systems* .

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

Cheng, H. T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Chai, W., Ispir, M., Anil, R., Haque, Z., Hong, L., Jain, V., Liu, X., & Shah, H. (2016). Wide & deep learning for recommender systems. *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, 7–10. https://doi.org/10.1145/2988450.2988454

Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine learning for pattern discovery in management research. *Strategic Management Journal*, *42*(1), 30–57. https://doi.org/10.1002/smj.3215

Chouldechova, A. (2017). Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big Data*, *5*(2), 153–163. https://doi.org/10.1089/big.2016.0047

Clarke, R. (2016). Big data, big risks. *Information Systems Journal*, *26*(1), 77–90. https://doi.org/10.1111/isj.12088

Coombs, C., Hislop, D., Taneva, S. K., & Barnard, S. (2020). The strategic impacts of Intelligent Automation for knowledge and service work: An interdisciplinary review. *Journal of Strategic Information Systems*, *29*(4), 1–30. https://doi.org/10.1016/j.jsis.2020.101600

Corbett-Davies, S., & Goel, S. (2018). The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. *ArXiv*. http://arxiv.org/abs/1808.00023

Daniel, W. W. (1990). *Applied Nonparametric Statistics* (2nd ed.). PWS-Kent Pub.

Daugherty, P. R., & Wilson, H. J. (2018). *Human + Machine: Reimagining Work in the Age of AI*. Harvard Business Review Press.

Davenport, T. H., & Kirby, J. (2016). *Only Humans Need Apply: Winners and Losers in the Age of Smart Machines*. HarperCollins.

de Bruyn, A., Viswanathan, V., Beh, Y. S., Brock, J. K. U., & von Wangenheim, F. (2020). Artificial Intelligence and Marketing: Pitfalls and Opportunities. *Journal of Interactive Marketing*, *51*(1), 91–105. https://doi.org/10.1016/j.intmar.2020.04.007

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

Delbecq, A., & van de Ven, A. (1971). A Group Process Model for Problem Identification and Program Planning. *Journal of Applied Behavioral Sci.*, *7*(4), 466–492. https://doi.org/10.1177/002188637100700404

Dempster, M. A. H., & Leemans, V. (2006). An automated FX trading system using adaptive reinforcement learning. *Expert Systems with Applications*, *30*(3), 543–552. https://doi.org/10.1016/j.eswa.2005.10.012

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

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

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

Dolata, M., Feuerriegel, S., & Schwabe, G. (2022). A sociotechnical view of algorithmic fairness. *Information Systems Journal*, *32*(4), 754–818. https://doi.org/10.1111/isj.12370

Domingos, P. (2012). A Few Useful Things to Know About Machine Learning. *Communications of the ACM*, *55*(10), 78–87. https://doi.org/10.1007/978-3-642-35289-8-3

Dougherty, D. (2016). *Taking Advantage of Emergence: Productively Innovating in Complex Innovation Systems*. Oxford University Press.

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

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, *28*(1), 62–70. https://doi.org/10.1016/j.infoandorg.2018.02.005

Farjoun, M., Ansell, C., & Boin, A. (2015). Pragmatism in organization studies: Meeting the challenges of a dynamic and complex world. *Organization Science*, *26*(6), 1787–1804. https://doi.org/10.1287/orsc.2015.1016

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 (accessed on Oct 30, 2022)

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

Flick, U. (2004). Triangulation in qualitative research. In *A companion to qualitative research* (3rd ed., pp. 173–183). Sage Publications.

Forbes Insights. (2018). *On Your Marks: Business Leaders Prepare For Arms Race In Artificial Intelligence*. Forbes. https://www.forbes.com/sites/insights-intelai/2018/07/17/on-your-marks-business-leaders-prepare-for-arms-race-in-artificial-intelligence/#4084b00f1946 (accessed on Oct 30, 2022)

Fu, R., Aseri, M., Singh, P., & Srinivasan, K. (2022). "Un"Fair Machine Learning Algorithms. *Management Science*, *68*(6), 3975–4753. https://doi.org/10.1287/mnsc.2021.4065

Fügener, A., Grahl, J., Gupta, A., & Keter, W. (2021). Cognitive challenges in human-AI collaboration: Investigating the path towards productive delegation. *Information Systems Research*, *33*(2), 678–696. https://doi.org/10.1287/isre.2021.1079

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Will Humans-in-the-Loop Become Borgs? Merits and Pitfalls of Working With AI. *MIS Quarterly*, *45*(3), 1527–1556. https://doi.org/10.25300/MISQ/2021/16553

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

Gavetti, G., & Levinthal, D. (2000). Looking Forward and Looking Backward: Cognitive and Experiential Search. *Administrative Science Quarterly*, *45*(1), 113–137. https://doi.org/10.2307/2666981

Gawlitza, J., Sturm, T., Spohrer, K., Henzler, T., Akin, I., Schönberg, S., Borggrefe, M., Haubenreisser, H., & Trinkmann, F. (2019). Predicting Pulmonary Function Testing from Quantified Computed Tomography Using Machine Learning Algorithms in Patients with COPD. *Diagnostics*, *9*(33), 1–13. https://doi.org/10.3390/diagnostics9010033

Gebauer, J., & Ginsburg, M. (2009). Exploring the black box of task-technology fit. *Communications of the ACM*, *52*(1), 130–135. https://doi.org/10.1145/1435417.1435447

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

Ghasemaghaei, M., Ebrahimi, S., & Hassanein, K. (2018). Data analytics competency for improving firm decision making performance. *Journal of Strategic Information Systems*, *27*(1), 101–113. https://doi.org/10.1016/j.jsis.2017.10.001

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

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

Grant, R. M. (1996). Toward a Knowledge-based Theory of the Firm. *Strategic Management Journal*, *17*(S2), 109–122. https://doi.org/10.1002/smj.4250171110

Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *Journal of Strategic Information Systems*, *29*(2), 1–16. https://doi.org/10.1016/j.jsis.2020.101614

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, *51*(5). https://doi.org/10.1145/3236009

Gupta, A. K., Smith, K. G., & Shalley, C. E. (2006). The Interplay Between Exploration And Exploitation. *Academy of Management Journal*, *49*(4), 693–706. https://doi.org/10.5465/amj.2006.22083026

Haas, C. (2019). The Price of Fairness - A Framework to Explore Trade-Offs in Algorithmic Fairness. *Proceedings of the 40th International Conference on Information Systems*.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the 2015 IEEE International Conference on Computer Vision*, 1026–1034. https://doi.org/10.1109/ICCV.2015.123

Heess, N., TB, D., Sriram, S., Lemmon, J., Merel, J., Wayne, G., Tassa, Y., Erez, T., Wang, Z., Eslami, S. M. A., Riedmiller, M., & Silver, D. (2017). Emergence of Locomotion Behaviours in Rich Environments. *ArXiv*, 1–14. http://arxiv.org/abs/1707.02286

Hippel, E. von, & Krogh, G. von. (2016). Problem solving without problem formulation. *Organization Science*, *27*(1), 207–221. https://doi.org/10.1287/orsc.2015.1023

Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Huber, G. P. (1991). Organizational Learning: The Contributing Processes and the Literatures. *Organization Science*, *2*(1), 88–115. https://doi.org/10.1287/orsc.2.1.88

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data Clustering: A Review. *ACM Computing Surveys*, *31*(3), 264–323. https://doi.org/10.1145/331499.331504

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa841

Junglas, I., Abraham, C., & Watson, R. T. (2008). Task-technology fit for mobile locatable information systems. *Decision Support Systems*, *45*(4), 1046–1057. https://doi.org/10.1016/j.dss.2008.02.007

Jussupow, E., Spohrer, K., Heinzl, A., & Gawlitza, J. (2021). Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. *Information Systems Research*, *32*(3), 713–735. https://doi.org/10.1287/ISRE.2020.0980

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

Kane, G. C., Young, A. G., Majchrzak, A., & Ransbotham, S. (2021). Avoiding an oppressive future of machine learning: A design theory for emancipatory assistants. *MIS Quarterly*, *45*(1), 371–396. https://doi.org/10.25300/MISQ/2021/1578

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling Laws for Neural Language Models. *ArXiv*. http://arxiv.org/abs/2001.08361

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), 109–213. https://doi.org/10.1287/isre.1040.0022

Kayser, V., Nehrke, B., & Zubovic, D. (2018). Data Science as an Innovation Challenge: From Big Data to Value Proposition. *Technology Innovation Management Review*, *8*(3), 16–25. https://doi.org/10.22215/timreview/1143

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, *14*(1), 366–410. https://doi.org/10.5465/annals.2018.0174

Kendall, M. G. (1945). The Treatment of Ties in Ranking Problems. *Biometrika*, *33*(3), 239–251. https://doi.org/10.2307/2332303

Kissel, R. L. (2013). *The Science of Algorithmic Trading and Portfolio Management*. Academic Press.

Kohli, R., & Melville, N. P. (2019). Digital innovation: A review and synthesis. *Information Systems Journal*, *29*(1), 200–223. https://doi.org/10.1111/isj.12193

Kolb, D. A. (1976). Management and the Learning Process. *California Management Review*, *18*(3), 21–31. https://doi.org/10.2307/41164649

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

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

Krupitzer, C., Drechsel, G., Mateja, D., Pollkläsener, A., Schrage, F., Sturm, T., Tomasovic, A., & Becker, C. (2018). Using Spreadsheet-defined Rules for Reasoning in Self-Adaptive Systems. *Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 289–294. https://doi.org/10.1109/PERCOMW.2018.8480283

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). AI for the Financial Services Industry: What Challenges Organizations to Succeed. *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Kumar, N., Venugopal, D., Qiu, L., & Kumar, S. (2018). Detecting Review Manipulation on Online Platforms with Hierarchical Supervised Learning. *Journal of Management Information Systems*, *35*(1), 350–380. https://doi.org/10.1080/07421222.2018.1440758

Kurgan, L. A., & Musilek, P. (2006). A survey of knowledge discovery and data mining process models. *Knowledge Engineering Review*, *21*(1), 1–24. https://doi.org/10.1017/S0269888906000737

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2020). ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. *Proceedings of the 8th International Conference on Learning Representations*, 1–17.

Lang, J. R., Dittrich, J. E., & White, S. E. (1978). Managerial Problem Solving Models: A Review and a Proposal. *Academy of Management Review*, *3*(4), 854–866. https://doi.org/10.5465/amr.1978.4289296

Lavie, D., Stettner, U., & Tushman, M. L. (2010). Exploration and Exploitation Within and Across Organizations. *Academy of Management Annals*, *4*(1), 109–155. https://doi.org/10.1080/19416521003691287

Leavitt, H. J. (1975). Beyond the Analytic Manager. *California Management Review*, *17*(3), 5–12. https://doi.org/10.2307/41162143

Leavitt, K., Schabram, K., Hariharan, P., & Barnes, C. M. (2021). Ghost in the Machine: On Organizational Theory in the Age of Machine Learning. *Academy of Management Review*, *46*(4), 750–777. https://doi.org/10.5465/amr.2019.0247

Lebovitz, S., Levina, N., & Lifshitz-Assaf, H. (2021). Is AI ground truth really "true"? The dangers of training and evaluating AI tools based on experts' know-what. *MIS Quarterly*, *45*(3), 1501–1526. https://doi.org/10.25300/MISQ/2021/16564

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. https://doi.org/10.1016/j.dss.2005.05.008

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

Levinthal, D. A. (1997). Adaptation on rugged landscapes. *Management Science*, *43*(7), 934–950. https://doi.org/10.1287/mnsc.43.7.934

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

Levinthal, D. A., & Rerup, C. (2021). The Plural of Goal: Learning in a World of Ambiguity. *Organization Science*, *32*(3), 527–543. https://doi.org/10.1287/orsc.2020.1383

Levitt, B., & March, J. G. (1988). Organizational Learning. *Annual Review of Sociology*, *14*(1), 319–338. https://doi.org/10.5465/amr.1985.4279103

Liebman, E., Saar-Tsechansky, M., & Stone, P. (2019). The Right Music At The Right Time: Adaptive Personalized Playlists Based On Sequence Modeling. *MIS Quarterly*, *43*(3), 765–786. https://doi.org/10.25300/MISQ/2019/14750

Lindberg, A. (2020). Developing theory through integrating human and machine pattern recognition. *Journal of the Association for Information Systems*, *21*(1), 90–116. https://doi.org/10.17705/1jais.00593

Lindebaum, D., Vesa, M., & den Hond, F. (2020). Insights from “the machine stops” to better understand rational assumptions in algorithmic decision making and its implications for organizations. *Academy of Management Review*, *45*(1), 247–263. https://doi.org/10.5465/amr.2018.0181

Lyles, M. A., & Fiol, C. M. (1985). Organizational Learning. *Academy of Management Review*, *10*(4), 803–813. https://doi.org/10.2307/258048

Lyles, M. A., & Mitroff, I. I. (1980). Organizational Problem Formulation: An Empirical Study. *Administrative Science Quarterly*, *25*(1), 102–119. https://doi.org/10.2307/2392229

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

Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, *18*(1), 50–60. https://doi.org/10.1214/aoms/1177730491

March, J. G. (1981). Footnotes to Organizational Change. *Administrative Science Quarterly*, *26*(4), 563–577. https://doi.org/10.1007/s13398-014-0173-7.2

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

March, J. G. (1994). *A Primer on Decision Making: How Decisions Happen*. Free Press.

March, J. G. (2006). Rationality, foolishness, and adaptive intelligence. *Strategic Management Journal*, *27*(3), 201–214. https://doi.org/10.1002/smj.515

March, J. G. (2010). *The Ambiguities of Experience*. Cornell University Press.

March, J. G., Sproull, L. S., & Tamuz, M. (1991). Learning from samples of one or fewer. *Organization Science*, *2*(1), 1–13. https://doi.org/10.1287/orsc.2.1.1

Marjanovic, O., Cecez-Kecmanovic, D., & Vidgen, R. (2022). Theorising Algorithmic Justice. *European Journal of Information Systems*, *31*(3), 269–287. https://doi.org/10.1080/0960085X.2021.1934130

Martin, K. (2019). Designing ethical algorithms. *MIS Quarterly Executive*, *18*(2), 129–142. https://doi.org/10.17705/2msqe.00012

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

Mateja, D., & Heinzl, A. (2021). Towards Machine Learning as an Enabler of Computational Creativity. *IEEE Transactions on Artificial Intelligence*, *2*(6), 460–475. https://doi.org/10.1109/TAI.2021.3100456

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence. *AI Magazine*, *27*(4), 12–14.

McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. C., 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

Michels, J., Saxena, A., & Ng, A. Y. (2005). High Speed Obstacle Avoidance Using Monocular Vision and Reinforcement Learning. *Proceedings of the 22nd International Conference on Machine Learning (ICML)*, 593–600. https://doi.org/10.1145/1102351.1102426

Miles, M. B., Huberman, A. M., & Saldana, J. (2013). *Qualitative Data Analysis: A Methods Sourcebook* (Vol. 3). SAGE.

Miller, G. A. (1956). The Magic Number Seven Plus or minus Two: Some Limits on Our Capacity for Processing Information. *The Psychology Review*, *63*(2), 81–97. https://doi.org/10.1177/001088049003100202

Miller, K. D., & Lin, S. J. (2010). Different Truths in Different Worlds. *Organization Science*, *21*(1), 97–114. https://doi.org/10.1287/orsc.1080.0409

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, T. M. (1997). *Machine Learning*. McGraw-Hill.

Mumford, E. (1998). Problems, knowledge, solutions: Solving complex problems. *Journal of Strategic Information Systems*, *7*, 255–269. https://doi.org/10.1016/S0963-8687(99)00003-7

Mumford, M. D., Whetzel, D. L., & Reiter-Palmon, R. (1997). Thinking creatively at work: Organization influences on creative problem solving. *Journal of Creative Behavior*, *31*(1), 7–17. https://doi.org/10.1002/j.2162-6057.1997.tb00777.x

Mumford, M., Reiter-Palmon, R., & Redmond, M. (1994). Problem construction and cognition: Applying problem representations in ill-defined domains. In M. A. Runco (Ed.), *Creativity research. Problem finding, problem solving, and creativity* (pp. 3–39). Ablex Publishing. https://psycnet.apa.org/record/1994-98186-001

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/doi.org/10.5465/amr.2019.0186

Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization*, *17*(1), 2–26. https://doi.org/10.1016/j.infoandorg.2006.11.001

Nickerson, J. A., & Zenger, T. R. (2004). A knowledge-based theory of the firm - The problem-solving perspective. *Organization Science*, *15*(6), 617–632. https://doi.org/10.1287/orsc.1040.0093

Nilsson, N. J. (1998). *Artificial Intelligence: A New Synthesis*. Morgan Kaufmann Publishers.

Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization Science*, *5*(1), 14–37. https://doi.org/10.1287/orsc.5.1.14

Osborn, A. F. (1953). *Applied imagination*. Charles Scribner's Sons.

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

Pentland, B. T., Recker, J., Wolf, J. R., & Wyner, G. (2020). Bringing context inside process research with digital trace data. *Journal of the Association for Information Systems*, *21*(5), 1214–1236. https://doi.org/10.17705/1jais.00635

Peters, F., Pumplun, L., & Buxmann, P. (2020). Opening the Black Box: Consumer's Willingness to Pay for Transparency of Intelligent Systems. *Proceedings of the 28th European Conference on Information Systems*.

Poole, D., Mackworth, A., & Goebel, R. (1998). *Computational Intelligence: A Logical Approach*. Oxford University Press.

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), 1–11. https://doi.org/10.1038/s41598-020-60191-3

Pumplun, L., Tauchert, C., & Heidt, M. (2019). A New Organizational Chassis for Artificial Intelligence - Exploring Organizational Readiness Factors. *Proceedings of the 27th European Conference on Information Systems (ECIS)*. https://aisel.aisnet.org/ecis2019_rp/106

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. (2021). Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review*, *46*(1), 192–210. https://doi.org/10.5465/2018.0072

Ransbotham, S., Khodabandeh, S., Kiron, D., Candelon, F., Chu, M., & LaFountain, B. (2020). Expanding AI's Impact With Organizational Learning. *MIT Sloan Management Review*, *62270*, 1–15. https://sloanreview.mit.edu/ai2020

Reiter-Palmon, R., & Robinson, E. J. (2009). Problem Identification and Construction: What Do We Know, What Is the Future? *Psychology of Aesthetics, Creativity, and the Arts*, *3*(1), 43–47. https://doi.org/10.1037/a0014629

Rhue, L. (2019). Beauty is in the AI of the Beholder: How Artificial Intelligence Anchors Human Decisions on Subjective vs. Objective Measures. *Proceedings of the 40th International Conference on Information Systems*.

Russell, S. J., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson Education.

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*.

Saldana, J. (2009). *The Coding Manual for Qualitative Researchers*. SAGE.

Salovaara, A., Lyytinen, K., & Penttinen, E. (2019). High Reliability in Digital Organizing: Mindlessness, the Frame Problem, and Digital Operations. *MIS Quarterly*, *43*(2), 555–578. https://doi.org/10.25300/MISQ/2019/14577

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

Sarker, S., & Valacich, J. S. (2010). An Alternative to Methodological Individualism: A Non-Reductionist Approach to Studying Technology Adoption by Groups. *MIS Quarterly*, *34*(4), 779–808. https://doi.org/10.2307/25750705

Sarker, S., Xiao, X., & Beaulieu, T. (2013). Guest Editorial: Qualitative Studies in Information Systems: A Critical Review and Some Guiding Principles. *MIS Quarterly*, *37*(4), iii–xviii. https://www.jstor.org/stable/43825778

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 (accessed on Oct 30, 2022)

Schilling, M. A., & Fang, C. (2014). When Hubs Forget, Lie, and Play Favorites: Interpersonal Network Structure, Information Distortion, and Organzational Learning. *Strategic Management Journal*, *35*(7), 974–994. https://doi.org/10.1002/smj

Schmelzer, R. (2019). *The AI-enabled Future*. Forbes. https://www.forbes.com/sites/cognitiveworld/2019/10/17/the-ai-enabled-future/#1260a9a53339 (accessed on Oct 30, 2022)

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

Searle, J. R. (1980). Minds, Brains, and Programs. *The Behavioral and Brain Sciences*, *3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G. J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information and Management*, *57*(2), 103174. https://doi.org/10.1016/j.im.2019.103174

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

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. (1972). Theories of Bounded Rationality. In C. B. McGuire & R. Radner (Eds.), *Decision and Organization* (pp. 161–176). Elsevier.

Simon, H. A. (1973). The structure of ill structured problems. *Artificial Intelligence*, *4*(3–4), 181–201. https://doi.org/https://doi.org/10.1016/0004-3702(73)90011-8

Simon, H. A. (1991). Bounded Rationality and Organizational Learning. *Organization Science*, *2*(1), 125–134. https://doi.org/10.1162/152417399570142

Simon, H. A., & Newell, A. (1971). Human Problem Solving: The State of the Theory in 1970. *American Psychologist*, *26*(2), 145–159. https://psycnet.apa.org/record/1971-24266-001

Smith, G. F. (1989). Defining Managerial Problems: A Framework for Prescriptive Theorizing. *Management Science*, *35*(8), 963–981. https://doi.org/10.1287/mnsc.35.8.963

Spearman, C. (1987). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, *100*(3/4), 441–471. https://doi.org/10.2307/1412159

Sturm, T., Fecho, M., & Buxmann, P. (2021). To Use or Not to Use Artificial Intelligence? A Framework for the Ideation and Evaluation of Problems to Be Solved with Artificial

Intelligence. *Proceedings of the 54th Hawaii International Conference on System Sciences*, 206–215. https://hdl.handle.net/10125/70634

Sturm, T., Gerlach, J. P., Pumplun, L., Mesbah, N., Peters, F., Tauchert, C., Nan, N., & Buxmann, P. (2021). Coordinating Human and Machine Learning for Effective Organizational Learning. *MIS Quarterly*, *45*(3), 1581–1602. 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. *Proceedings of the 42nd International Conference on Information Systems*. https://aisel.aisnet.org/icis2021/ai_business/ai_business/3/

Sturm, T., Krupitzer, C., Segata, M., & Becker, C. (2021). A Taxonomy of Optimization Factors for Platooning. *IEEE Transactions on Intelligent Transportation Systems*, *22*(10), 6097–6114. https://doi.org/10.1109/tits.2020.2994537

Sturm, T., & Peters, F. (2020). The Impact of Artificial Intelligence on Individual Performance: Exploring the Fit between Task, Data, and Technology. *Proceedings of the 28th European Conference on Information Systems*. https://aisel.aisnet.org/ecis2020_rp/200/

Taylor, R. N. (1975). Perception of Problem Constraints. *Management Science*, *22*(1), 22–29. https://doi.org/10.1287/mnsc.22.1.22

Teodorescu, M. H. M., Morse, L., Awwad, Y., & Kane, G. C. (2021). Failures of Fairness in Automation Require a Deeper Understanding of Human–ML Augmentation. *MIS Quarterly*, *45*(3), 1483–1500. https://doi.org/10.25300/MISQ/2021/16535

Togelius, J., Karakovskiy, S., Koutník, J., & Schmidhuber, J. (2009). Super Mario Evolution. *Proceedings of the 2009 IEEE Symposium on Computational Intelligence and Games (CIG)*, 156–161. https://doi.org/10.1109/CIG.2009.5286481

Traumer, F., Oeste-Reiß, S., & Leimeister, J. M. (2017). Towards a Future Reallocation of Work between Humans and Machines – Taxonomy of Tasks and Interaction Types in the Context of Machine Learning. *Proceedings of the 38th International Conference on Information Systems*. https://doi.org/10.2139/ssrn.3159131

Tremblay, M., Kohli, R., & Forsgren, N. (2021). Theories in Flux: Reimagining Theory Building in the Age of Machine Learning. *MIS Quarterly*, *45*(1), 455–459. https://doi.org/10.25300/MISQ/2021/15434.1

Tschang, F. T., & Almirall, E. (2021). Artificial intelligence as augmenting automation: Implications for employment. *Academy of Management Perspectives*, *35*(4), 642–659. https://doi.org/10.5465/amp.2019.0062

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *LIX*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

University of Cambridge. (2022). *Coordination*. Cambridge Advanced Learner's Dictionary & Thesaurus; Cambridge University Press. https://dictionary.cambridge.org/us/dictionary/english/coordination (accessed on Oct 30, 2022)

van den Broek, E., Sergeeva, A., & Huysman, M. (2019). Hiring Algorithms: An Ethnography of Fairness in Practice. *Proceedings of the 40th International Conference on Information Systems*.

van den Ende, J., Frederiksen, L., & Prencipe, A. (2015). The front end of innovation: Organizing search for ideas. *Journal of Product Innovation Management*, *32*(4), 482–487. https://doi.org/10.1111/jpim.12213

Vanauer, M., Bohle, C., & Hellingrath, B. (2015). Guiding the introduction of big data in organizations: A methodology with business- and data-driven ideation and enterprise architecture management-based implementation. *Proceedings of the 48th Hawaii International Conference on System Sciences*, *2015-March*, 908–917.

Venturini, T., & Latour, B. (2009). The Social Fabric: Digital Traces and Quali-Quantitative Methods. *Proceedings of Future En Seine*.

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

Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1–37. https://doi.org/10.2478/jagi-2019-0002

Watkins, C. J. C. H., & Dayan, P. (1992). Q-Learning. *Machine Learning*, *8*, 55–68. https://doi.org/10.1109/ICCC49849.2020.9238991

Weber, R. P. (1990). *Basic content analysis*. Sage Publications.

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization Science*, *16*(4), 409–421. https://doi.org/10.1287/orsc.1050.0133

Williamson, O. E. (1981). The Economics of Organization: The Transaction Cost Approach. *American Journal of Sociology*, *87*(3), 548–577. https://doi.org/10.1086/227496

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Application of Knowledge Discovery and Data Mining*, 29–39. https://doi.org/10.1.1.198.5133

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. https://doi.org/10.1108/02686900010344511

Yin, R. K. (2011). *Applications of Case Study Research* (3rd ed.). SAGE Publications.

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