Adoption of AI-based Information Systems from an Organizational and User Perspective

Vom Fachbereich Rechts- und Wirtschaftswissenschaften
der Technischen Universität Darmstadt

genehmigte

Dissertation

von

Christoph Tauchert, M.Sc.
geboren in Gießen

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

Erstgutachter:	Prof. Dr. Peter Buxmann
Zweitgutachter:	Prof. Dr. Alexander Benlian
Darmstadt	2022

Tauchert, Christoph: Adoption of AI-based Information Systems from an Organizational and User Perspective

Darmstadt, Technische Universität Darmstadt

Dissertation veröffentlicht auf TUprints im Jahr 2022
Tag der mündlichen Prüfung: 17.11.2022

Declaration of Authorship

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

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

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

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

Christoph Tauchert

Darmstadt, 18.05.2021

Abstract

Artificial intelligence (AI) is fundamentally changing our society and economy. Companies are investing a great deal of money and time into building corresponding competences and developing prototypes with the aim of integrating AI into their products and services, as well as enriching and improving their internal business processes. This inevitably brings corporate and private users into contact with a new technology that functions fundamentally differently than traditional software. The possibility of using machine learning to generate precise models based on large amounts of data capable of recognizing patterns within that data holds great economic and social potential—for example, in task augmentation and automation, medical diagnostics, and the development of pharmaceutical drugs. At the same time, companies and users are facing new challenges that accompany the introduction of this technology. Businesses are struggling to manage and generate value from big data, and employees fear increasing automation. To better prepare society for the growing market penetration of AI-based information systems into everyday life, a deeper understanding of this technology in terms of organizational and individual use is needed.

Motivated by the many new challenges and questions for theory and practice that arise from AI-based information systems, this dissertation addresses various research questions with regard to the use of such information systems from both user and organizational perspectives. A total of five studies were conducted and published: two from the perspective of organizations and three among users. The results of these studies contribute to the current state of research and provide a basis for future studies. In addition, the gained insights enable recommendations to be derived for companies wishing to integrate AI into their products, services, or business processes.

The first research article (Research Paper A) investigated which factors and prerequisites influence the success of the introduction and adoption of AI. Using the technology–organization–environment framework, various factors in the categories of technology, organization, and environment were identified and validated through the analysis of expert interviews with managers experienced in the field of AI. The results show that factors related to data (especially availability and quality) and the management of AI projects (especially project management and use cases) have been added to the framework, but regulatory factors have also emerged, such as the uncertainty caused by the General Data Protection Regulation.

The focus of Research Paper B is companies' motivation to host data science competitions on online platforms and which factors influence their success. Extant research has shown that employees with new skills are needed to carry out AI projects and that many companies have problems recruiting such employees. Therefore, data science competitions could support the implementation of AI projects via crowdsourcing. The results of the study (expert interviews among data scientists) show that these competitions offer many advantages, such as exchanges and discussions with experienced data scientists and the use of state-of-the-art approaches. However, only a small part of the effort related to AI projects can be represented within the framework of such competitions.

The studies in the other three research papers (Research Papers C, D, and E) examine AI-based information systems from a user perspective, with two studies examining user behavior and one focusing on the design of an AI-based IT artifact. Research Paper C analyses perceptions of AI-based advisory systems in terms of the advantages associated with their use. The results of the empirical study show that the greatest perceived benefit is the convenience such systems provide, as they are easy to access at any time and can immediately satisfy informational needs. Furthermore, this study examined the effectiveness of 11 different measures to increase trust in AI-based advisory systems. This showed a clear ranking of measures, with effectiveness decreasing from non-binding testing to providing additional information regarding how the system works to adding anthropomorphic features.

The goal of Research Paper D was to investigate actual user behavior when interacting with AI-based advisory systems. Based on the theoretical foundations of task–technology fit and judge–advisor systems, an online experiment was conducted. The results show that, above all, perceived expertise and the ability to make efficient decisions through AI-based advisory systems influence whether users assess these systems as suitable for supporting certain tasks. In addition, the study provides initial indications that users might be more willing to follow the advice of AI-based systems than that of human advisors.

Finally, Research Paper E designs and implements an IT artifact that uses machine learning techniques to support structured literature reviews. Following the approach of design science research, an artifact was iteratively developed that can automatically download research articles from various databases and analyze and group them according to their content using the word2vec algorithm, the latent Dirichlet allocation model, and agglomerative hierarchical cluster analysis. An evaluation of the artifact on a dataset of 308 publications shows that it can be a helpful tool to support literature reviews but that much manual effort is still required, especially with regard to the identification of common concepts in extant literature.

Zusammenfassung

Künstliche Intelligenz (KI) ist gerade dabei, unsere Gesellschaft und Wirtschaft fundamental zu verändern. Unternehmen investieren große Geldsummen in den Aufbau entsprechender Kompetenzen und die Entwicklung von Prototypen, mit dem Ziel, KI in ihre Produkte und Dienstleistungen zu integrieren oder aber auch ihre internen Unternehmensprozesse damit anzureichern und zu verbessern. Hierdurch haben zwangsläufig Unternehmens- aber auch Privatanwender Kontakt mit einer Technologie, welche im Kern anders funktioniert als traditionelle Software. Die Möglichkeit, durch maschinelles Lernen auf der Basis großer Datenmengen präzise Modelle zu erzeugen, die in der Lage sind, Muster in diesen Daten zu erkennen, birgt große ökonomische und gesellschaftliche Potenziale – beispielsweise in der Augmentation und Automatisierung von Aufgaben aber auch in der medizinischen Diagnostik und der Entwicklung von Medikamenten. Allerdings stehen Unternehmen und Anwender gleichzeitig auch vor neuen Herausforderungen, die mit der Einführung dieser Technologie einhergehen. So kämpfen Unternehmen unter anderem damit, die großen Datenmengen zu verwalten und Werte daraus zu generieren. Gleichzeitig fürchten Angestellte um ihre Jobs durch die zunehmende Automatisierung. Um unsere Gesellschaft besser auf die wachsende Durchdringung des Alltags durch KI-basierte Informationssysteme vorzubereiten, wird ein tiefergehendes Verständnis dieser Technologie hinsichtlich organisationaler und individueller Nutzung benötigt.

Motiviert durch die vielen neuen Herausforderungen und Fragestellungen für Theorie und Praxis, die sich durch KI-basierte Informationssysteme ergeben, werden in dieser Arbeit verschiedene Forschungsfragen im Hinblick auf die Nutzung solcher Informationssysteme adressiert – sowohl aus Anwenderperspektive als auch aus organisationaler Perspektive. Insgesamt wurden fünf Studien durchgeführt und publiziert: zwei Studien aus der Perspektive von Unternehmen sowie drei Studien unter Anwendern. Die Ergebnisse der durchgeführten Untersuchungen tragen zum aktuellen Stand der Forschung bei und stellen eine Basis für zukünftige Studien dar. Zudem ermöglichen die herausgearbeiteten Erkenntnisse das Ableiten von Handlungsempfehlungen für Unternehmen, welche KI in ihre Produkte, Dienste oder Unternehmensprozesse integrieren wollen.

Im Rahmen des ersten Forschungsartikels (Forschungspapier A) wurde untersucht, welche Faktoren und Voraussetzungen den Erfolg bei der Einführung und Annahme von künstlicher Intelligenz beeinflussen. Basierend auf dem „Technologie-Organisation-Environment-Framework“ wurden durch die Analyse von Experteninterviews mit Managern, die Erfahrungen im Bereich KI gesammelt haben, verschiedene Faktoren in den Kategorien Technologie, Organisation und Umwelt identifiziert und validiert. Die Ergebnisse zeigen, dass vor allem Faktoren in Zusammenhang mit den Daten (insb. Verfügbarkeit, Qualität) sowie im Umgang mit KI-Projekten (insb. Projektmanagement, Anwendungsfälle) hinzugekommen sind, aber auch regulatorische Faktoren wie die Unsicherheit durch die Datenschutzgrundverordnung zu berücksichtigen sind.
Der Fokus von Forschungspapier B liegt darin, zu untersuchen, aus welchen Beweggründen Unternehmen Data Science Wettbewerbe auf Online Plattformen veranstalten und welche Faktoren deren Erfolg beeinflussen. Bestehende Forschung hat gezeigt, dass zur Durchführung von KI-Projekten Mitarbeiter mit neuen Fähigkeiten gebraucht werden und das viele Unternehmen Probleme haben, entsprechende Mitarbeiter zu rekrutieren. Daher könnten solche Wettbewerbe eine Möglichkeit sein, die Durchführung von KI-Projekten per Crowdsourcing zu unterstützen. Die Ergebnisse der Studie (Experteninterviews) unter Datenwissenschaftlern zeigen, dass die Wettbewerbe viele Vorzüge bieten, wie bspw. der Austausch mit erfahrenen Datenwissenschaftlern sowie die Verwendung neuster Ansätze. Allerdings kann nur ein kleiner Teil des Aufwands, der bei KI-Projekten anfällt, im Rahmen solcher Wettbewerbe abgebildet werden.
Die Studien der anderen drei Forschungsartikel (Forschungspapier C, D und E) betrachten KI-basierte Informationssysteme aus Anwenderperspektive – wobei zwei Studien das Verhalten von Nutzern untersuchen und bei einer Studie die Erstellung eines KI-basierten IT-Artefakts im Fokus steht. Forschungspapier C analysiert die Wahrnehmung von KI-basierten Beratungssystemen hinsichtlich der Vorteile, die sich aus deren Nutzung ergeben. Die Ergebnisse der empirischen Studie zeigen, dass der größte wahrgenommene Nutzen der gebotene Komfort ist, da die Systeme einfach und jederzeit zugänglich sind sowie eine sofortige Befriedigung des Informationsbedürfnisses bieten. Des Weiteren, wurde in dieser Studie die Wirksamkeit elf verschiedener Maßnahmen zur Steigerung des Vertrauens in KI-basierte Beratungssysteme untersucht. Hier zeigte sich eine klare Rangfolge der Maßnahmen mit abnehmender Wirksamkeit vom unverbindlichen Testen über das Bereitstellen zusätzlicher Informationen bezüglich der Funktionsweise des Systems hin zu dessen Vermenschlichung.

Das Ziel von Forschungspapier D ist die Untersuchung des tatsächlichen Nutzungsverhalten von Anwendern während der Interaktion mit KI-basierten Beratungssystemen. Basierend auf den theoretischen Grundlagen des „Task-Technology Fit“ und von „Judge-Advisor Systems“ wurde ein Onlineexperiment durchgeführt. Die Ergebnisse zeigen, dass vor allem die wahrgenommene Expertise als auch die Möglichkeit, durch KI-basierte Beratungssysteme effizient Entscheidungen treffen zu können, beeinflussen, ob Anwender diese Systeme als geeignet zur Unterstützung bestimmter Aufgaben einschätzen. Zudem gibt die Studie erste Hinweise darauf, dass Anwender den Ratschlägen KI-basierter Systeme mehr Beachtung schenken könnten als denen menschlicher Berater.

Schließlich wird in Forschungspapier E ein IT-Artefakt entworfen und implementiert, welches durch Nutzung von Techniken des maschinellen Lernens das Durchführen von strukturierten Literaturrecherchen unterstützt. Dem Vorgehen der „Design Science“ Forschung folgend, wurde iterativ ein Artefakt entwickelt, welches automatisiert Forschungsartikel von verschiedenen Datenbanken herunterladen sowie diese hinsichtlich ihres Inhalts analysieren und gruppieren kann. Hierzu wurde insbesondere der „word2vec“ Algorithmus, das „Latent Dirichlet Allocation“ Modell und agglomerative hierarchische Clusteranalyse verwendet. Eine Bewertung des Artefakts auf einem Datensatz von 308 Veröffentlichungen zeigt, dass es ein hilfreiches Werkzeug zur Unterstützung bei Literaturrecherchen sein kann, aber insbesondere hinsichtlich der Identifikation gemeinsamer Konzepte weiterhin viel manueller Aufwand erforderlich ist.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AEX	Advisor’s Expertise
AI	Artificial Intelligence
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
API	Application Programming Interface
AVE	Average Variance Extracted
B2B	Business-to-Business
B2C	Business-to-Consumer
CEO	Chief Executive Officer
CIO	Chief Information Officer
CR	Composite Reliability
Cr. α	Cronbach's α
CRISP-DM	Cross Industry Standard Process for Data Mining
CRM	Customer Relationship Management
EFF	Advisor’s Efficiency-Enhancing Characteristic
EMO	Emotional Trust in Advisor
GDPR	General Data Protection Regulation
HR	Human Resources
HU	Human Advisor
INT	Advisor Integrity
IS	Information System(s)
IT	Information Technology
JAS	Judge-Advisor System
KPI	Key Performance Indicator
LDA	Latent Dirichlet Allocation
LSA	Latent Semantic Analysis
OCR	Optical Character Recognition
PDF	Portable Document Format
PF	Provider Firm

pLSA	Probabilistic Latent Semantic Analysis
RA	Relative Advantage
RA	Robo-Advisor
RAIC	Relative Advantage - Informational Convenience
RAIT	Relative Advantage - Informational Trust
RAKE	Rapid Automatic Keyword Extraction
RAL	Relative Advantage - Learning
RAT	Relative Advantage - Transaction
RO	Research Objective
ROI	Return on Investment
SRMR	Standardized Root Mean Square Residual
TAF	Task-Advisor-Fit
TAM	Technology Acceptance Model
TOE	Technological-Organizational-Environmental
TTF	Task-Technology-Fit
UEX	User's Expertise
UF	User Firm
UTAUT	Unified Theory of Acceptance and Use of Technology
WOA	Weight of Advice

1 Introduction

> *The most important general-purpose technology of our era is artificial intelligence, particularly machine learning […]. The effects of AI will be magnified in the coming decade, as […] virtually every […] industry transform[s] [its] core processes and business models to take advantage of machine learning. The bottleneck now is in management, implementation, and business imagination.*

Erik Brynjolfsson and Andrew McAfee (2017)

1.1 Motivation

The term artificial intelligence (AI) was coined in a 1956 workshop at Dartmouth College by John McCarthy, who described AI as “the science and engineering of making intelligent machines, especially intelligent computer programs” (McCarthy 2007, p. 2). While the field of AI experienced several setbacks in the mid-1970s and late 1980s through early 1990s, research on AI—especially the advancement of deep learning techniques—has rapidly gained momentum since the early 2010s (e.g., Schwartz et al. 2020; Yoav Shoham et al. 2018). Additionally, the application of AI in organizations also started to gain momentum in the early 2010s, resulting in Gartner declaring AI a mega-trend and the most disruptive class of technology for the decade in 2017 (Panetta 2017). A core technology of AI is machine learning (ML), which describes the ability of computer programs to improve their performance at a given task using data (Mitchell 1997). To achieve this, complex approaches with varying degrees of transparency and explainability are applied. AI models are thus often referred to as “black boxes” (e.g., Adadi and Berrada 2018; Miller and Brown 2018), although some approaches—such as linear regression and decision trees—are inherently transparent and explainable (e.g., Letham et al. 2015).

Due to the ubiquity of information technology (IT) and sensors today, vast amounts of data are generated and stored continuously when using digital products or services (Statista 2020; Wanner et al. 2020). This data, along with the information that can be extracted from it, represents assets for the corresponding enterprises (Ackoff 1989; Chen et al. 2012). As more and more computational power becomes available and the cost of storing large amounts of data decreases, organizations can analyze this data to discover knowledge, support decisions, and even automate or augment tasks using AI (Brynjolfsson et al. 2011; McAfee and Brynjolfsson 2012). As a general-purpose technology, AI can be and has been applied to a variety of scenarios, including predictive maintenance of machines, drug discovery, and fraud detection (e.g., Awoyemi et al. 2017; Fleming 2018; Kanawaday and Sane 2018; Kwon et al. 2019). Furthermore, AI can be used to drive different types of analyses with regard to difficulty and value (e.g., Akerkar 2013; Krumeich et al. 2016; Lepenioti et al. 2020): (i) descriptive analytics to explain historical behavior, answering the questions “What has happened?” and “Why did it happen?”; (ii) predictive analytics to anticipate future events (“What will happen?” and “Why will it happen?”); and (iii) prescriptive analytics to provide advice on possible actions to take to achieve an optimal outcome (“What should I do?” and “Why should I do it?”). Due to the opportunities offered by information systems (IS) based on AI algorithms (hereafter, “AI-based IS”), such systems have received increasing attention from both technology companies and more “traditional” companies that anticipate related competitive advantages (MSV 2018).

However, organizations are struggling to realize the potential of AI and turn data into real value. A study by the Ericsson IndustryLab showed that almost every organization is facing difficulties in its journey to operationalize AI, and as many as 91% have to address challenges in the three areas—technology, people, and organization (Ericsson 2020). Remarkably, most companies are facing more challenges related to managing people and culture (e.g., change and innovation resistance, fears regarding loss of jobs and control) than technological challenges (e.g., need for dedicated hard- and software, data management). Not adequately addressing these challenges can result in failing AI projects during various phases, such as non-working prototypes or failure to operationalize and scale projects. According to Gartner research, only 53% of AI projects that succeed in developing a proof of concept also accomplish the transition from prototype to production (Gartner 2020). Even large technology companies are struggling with failing AI projects: Microsoft shut down its Twitter bot Tay after it posted racist slurs that it had learned from user interactions (Reese 2016), while IBM cancelled Watson for Oncology after years of development and spending USD 62 million (Strickland 2019). Nonetheless, successful AI projects show that perseverance is rewarded. A recent McKinsey study showed that organizations that master the challenges of the AI journey generate three to four times the returns from their investments (Atsmon et al. 2021).

Given the great potential benefits of AI and ML, even more organizations will begin to invest time and money in integrating AI into their services, products, and processes. 52% of US companies accelerated their AI investments even in the wake of the COVID-19 crisis (PwC 2021). Therefore, it is crucial to understand exactly what challenges organizations are facing in the context of this "new" technology, why these challenges arise, and how organizations can overcome them to eventually achieve the adoption and implementation of AI. To obtain a holistic picture of AI adoption, it is necessary to approach the research object from different perspectives. Besides understanding organizational challenges, it is necessary to investigate how users interact with IS with integrated AI capabilities. Additionally, research is needed to understand how AI-based IS can be designed to optimally support users in performing specific tasks.

From a theoretical point of view, the use of IS and the process of their adoption have been the focus of many studies in IS research (Burton-Jones et al. 2017). However, due to the newness of AI's importance to economy and society, research on the use of this technology is comparably scarce. Generally, it must be considered that decisions towards the adoption and use of a general-purpose technology such as AI is usually context-specific. Accordingly, existing studies analyzing the adoption of new technologies can primarily be divided into two categories: studies in the user context and studies in the organizational context.

Within the organizational context, extant research on AI is diverse. Some studies have focused on the implications on the future of work and the workforce (e.g., vom Brocke et al. 2018; Brynjolfsson and Mitchell 2017), while others have focused on specific industries, such as finance or service (e.g., Huang and Rust 2018; Kruse et al. 2019), and yet others have investigated how specific organizational processes can be technically supported by AI (e.g., Moncrief 2017). However, few studies have sought to provide a holistic picture of the organizational aspects of AI adoption and its implementation into organizational processes and governance structures (Ransbotham et al. 2017). Nonetheless, two studies have already, based on the extant literature, identified some aspects that should be considered when investigating AI adoption (e.g., commitment, technological expertise) (Alsheibani et al. 2018; Nascimento et al. 2018) as well as a possible research framework (Alsheibani et al. 2018). These studies can serve as a starting point for empirical studies to gain a better understanding of the factors that drive or impede the adoption of AI in organizations. Moreover, no existing study has investigated the capabilities of crowdsourcing ML models through data science communities.

This approach could potentially help organizations tackle challenges related to their lack of resources and experience related to AI.

From the user perspective, research on the acceptance and use of AI-based IS is also rare. Since AI-based chatbots and voice assistants (e.g., Siri, Alexa) were among the first AI-based products with which private users had intense contact, studies focused on investigating the differences between users' interactions with chatbots and those with their human counterparts. Conversational behavior was found to differ with regard to users' self-disclosure, message length, and content (e.g., Hill et al. 2015; Mou and Xu 2017; Pickard et al. 2016). Other studies focused on investigating the effects of different system characteristics, such as transparency or appearance, on user behavior and showed that transparency about the system's decision significantly influenced user behavior. For example, increased transparency positively affects users' perceptions of such a system, their satisfaction with its recommendations, and their decision-making effectiveness (e.g., Gedikli et al. 2014; Gregor 2001; Xu et al. 2014). Additionally, many studies have confirmed the positive effects of human-like appearances on user engagement and perceptions of trust and enjoyment (e.g., Qiu and Benbasat 2008; Schuetzler et al. 2018). However, while many characteristics that increase trust in AI-based systems have been identified, there is a lack of comparative studies evaluating their differences in efficacy. Furthermore, as AI is increasingly integrated into a variety of IS, it needs to be investigated how users perceive specific AI-based IS—such as AI-based advisory services—and their generated output. Many studies exploring the use of AI-based systems build on quantitative data collected via questionnaires. Therefore, analyses and results are based on adoption intentions as stated by the participants. However, psychological research shows that intentions do not necessarily lead to consistent actual behaviors (Sheeran 2002).

1.2 Research Objectives

Overall, within the user and organizational contexts, many research questions regarding the organizational adoption and individual use of AI remain unanswered, as the technology has just recently gained significant importance (e.g., Berente et al. 2019; Nascimento et al. 2018; Rai et al. 2019). Considering the rapid evolution of AI and the increasing opportunities for its application, the need for further research on organizational AI adoption as well as individuals' perceptions of AI-based IS is even more urgent. To address this gap, the overarching objective of this dissertation is to advance the understanding of the adoption of AI among organizations and users. This dissertation includes five studies, each addressing different research questions. The results of the five studies were published in the proceedings of various conferences and provide numerous theoretical contributions as well as practical implications for organizations. In the following, a summary of the five research objectives (ROs) is presented. A more detailed presentation of the motivation, research gaps, and derived research questions can be found in the chapters on the respective studies.

IS research has just recently begun to examine the organizational aspects of the adoption of AI. Therefore, few studies have dealt with general organizational aspects such as the implementation of AI in organizational processes and governance (e.g., Alsheibani et al. 2018; Baier et al. 2019; Ransbotham et al. 2017). While some published studies have focused on the impact of AI on specific business units and industries (e.g., Huang and Rust 2018; Kruse et al. 2019), a literature review by Nascimento et al. (2018) reveals that overarching mechanisms—such as the influence of an organization’s strategy or environment on the implementation of AI-based IS—have largely been neglected. Therefore, drawing on the technology–organization–environment (TOE) framework and innovation diffusion theory (DePietro et al. 1990; Rogers 2003; Zhu and Kraemer 2005), the study in chapter 3 seeks to provide holistic insights by exploring factors required for a successful adoption and implementation of AI in organizations.

RO 1: *Exploring readiness factors of organizational AI adoption.*

The aforementioned literature review also identified specific aspects that should be considered with regard to the adoption of AI, such as the specific skills employees must possess to handle AI technology (Nascimento et al. 2018). Additionally, practice-oriented research papers emphasize the need for new job profiles and the necessity of developing the skills of the workforce (e.g., vom Brocke et al. 2018). Lastly, practitioners themselves have noted the importance of correctly staffing AI teams and are worried about an acute shortage of AI talent (e.g., Costello 2019; Goasduff 2020). At the same time, many aspiring and experienced data scientists are participating in data science communities such as Kaggle and are challenging themselves by competing in ML competitions. Therefore, the study in chapter 4 seeks to investigate the capability of data science competitions to supplement organizations’ AI projects.

RO 2: *Analyzing organizations’ motives to host data science competitions and associated success factors.*

As mentioned in section 1.1, previous research on the introduction and use of new technologies has not only taken an organizational perspective but has also investigated users’ perceptions of and interactions with such technologies. Likewise, it is necessary to analyze and evaluate the behavior of individuals when using AI-based IS (e.g., Rzepka and Berger 2018). Analogous to the organizational context, due to the novelty of the technology, there are also many research gaps in the private context. Research is especially rare on users' perceptions of the advantages of AI-based IS, users' actual behavior while using AI-based IS, and the design of AI-based IS.

One area in which users are increasingly interacting with AI-based IS is automated advisory services that use AI to generate advice in a specific field of knowledge (e.g., wealth management, insurance) (e.g., Jung, Dorner, Weinhardt, et al. 2018; Kruse et al. 2019). However, these systems have not lived up to expectations in terms of their numbers of users (Jung and Weinhardt 2018). When it comes to AI-based advisory services, two concepts have received little attention but are key factors in adoption: first, the relative advantage that users perceive compared to other services (Choudhury and Karahanna 2008), and second, users' trust in the advisory service's advice (Lin 2011; Pavlou 2018). Previous research has focused on identifying opportunities to manipulate perceived trust in the system (e.g., Nilashi et al. 2016; de Visser et al. 2016). However, there is a need to compare different mechanisms and assess their relative effectiveness because their implementation is mostly not trivial and requires substantial resources. Therefore, the study in chapter 5 seeks to investigate users' perceptions of AI-based advisory services with the following research objective:

> **RO 3:** *Analyzing users' perceptions of the advantages of AI-based advisory systems and their trust in these systems.*

Additionally, extant research shows that people tend to reject help from IS—a phenomenon called algorithm aversion (e.g., Castelo et al. 2019; Dietvorst et al. 2015; Jussupow and Benbasat 2020). This phenomenon could further hinder customers' acceptance and use of AI-based advisory systems (e.g., financial robo-advisors). However, some studies have also found evidence of automation bias, wherein people tend to simply adopt an algorithmic recommendation (e.g., Skitka et al. 1999). To investigate this behavior further, the study in chapter 6 makes use of the judge–advisor system (JAS) paradigm that has been used in many studies in the cognitive sciences to analyze the behavior of individuals and groups when giving and taking advice with regard to various factors—such as judge and advisor characteristics or the type of interaction (e.g., Harvey and Fischer 1997; Sniezek and Van Swol 2001; Van Swol and Sniezek 2005). However, these studies have mostly investigated interactions between human actors, while the interplay between humans and IS has received less attention (Bonaccio and Dalal 2006). A drawback of many studies on the use of IS is that they measure only intention to use, rather than actual use (e.g., Hein et al. 2018; Lowry et al. 2012). This study seeks to investigate actual behavior by adopting experimental approaches from the JAS literature.

RO 4: *Analyzing users' actual advice-taking behavior with regard to perceived advisor characteristics.*

In IS research, design science is (besides behavioral science) one of the most widely used paradigms (Hevner et al. 2004). Among other objectives, it aims to create IT artifacts that extend the boundaries of human problem-solving by providing intellectual and computational tools (Hevner et al. 2004). The study in chapter 7 uses design science research to evaluate the possibilities of using ML to support the important process of structured scientific literature reviews. Extant research proposes many algorithms and techniques to analyze natural language and specifically for topic segmentation (e.g., Blei 2012; Griffiths and Steyvers 2004). However, these algorithms are usually developed and evaluated in isolation, whereas this study seeks to integrate ML approaches for text processing and topic segmentation into an end-to-end process for literature reviews.

RO 5: *Design of an IT artifact that uses ML to support the process of structured literature reviews.*

1.3 Dissertation Structure

This dissertation is organized into eight chapters (see Figure 1). Following the introduction, which lays out the motivation, ROs, and structure of the dissertation, the theoretical and methodological fundamentals are presented in chapter 2. Subsequently, five peer-reviewed and published research papers represent the core of this cumulative dissertation. The first two research papers focus on different aspects of AI from an organizational perspective (chapters 3 and 4). The other three focus on the user perspective of AI-based IS. While two investigate users' perceptions of AI-based advisory systems (chapters 5 and 6), the last focuses on designing an IT artifact that uses AI (chapter 7). Lastly, a brief summary of key findings, theoretical contributions, and practical implications is presented in chapter 8.

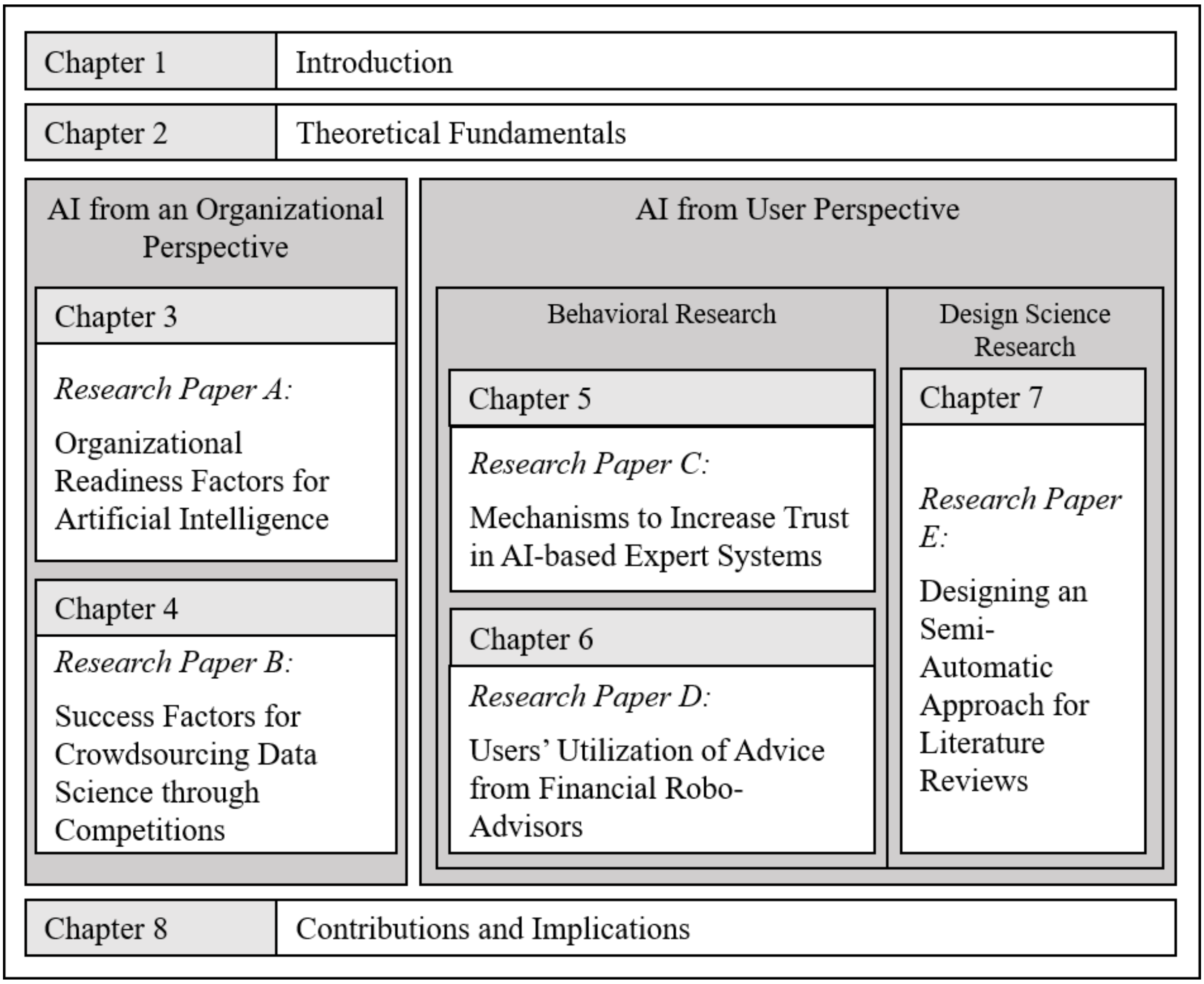

Figure 1. Dissertation Structure

In the following section, the five research papers included in this dissertation are listed and briefly summarized. As these studies were conducted with co-authors, the summaries and the research papers in chapters 3 through 7 also represent their work and opinions. The five research papers, along with a list of authors, publication outlets, publication years, and publication rankings, are listed in Table 1.

Paper A	Pumplun, Luisa; Tauchert, Christoph; Heidt, Margareta (2019): **A New Organizational Chassis for Artificial Intelligence – Exploring Organizational Readiness Factors.** In: European Conference on Information Systems (ECIS), Stockholm, Sweden. VHB-Ranking: B.
Paper B	Tauchert, Christoph; Buxmann, Peter; Lambinus, Jannis (2020): **Crowdsourcing Data Science: A Qualitative Analysis of Organizations' Usage of Kaggle Competitions.** In: Hawaii International Conference on System Sciences (HICSS), Wailea, Hawaii, USA. VHB-Ranking: C.

Paper C	Mesbah, Neda; Tauchert, Christoph; Olt, Christian M.; Buxmann, Peter (2019): **Promoting Trust in AI-Based Expert Systems.** In: Americas Conference on Information Systems (AMCIS), Cancun, Mexico. VHB-Ranking: D.
Paper D	Tauchert, Christoph; Mesbah, Neda (2019): **Following the Robot? Investigating Users' Utilization of Advice from Robo-Advisors.** In: International Conference on Information Systems (ICIS), Munich, Germany. VHB-Ranking: A.
Paper E	Tauchert, Christoph; Bender, Marco; Mesbah, Neda; Buxmann, Peter (2020): **Towards an Integrative Approach for Automated Literature Reviews Using Machine Learning.** In: Hawaii International Conference on System Sciences (HICSS), Wailea, Hawaii, USA. VHB-Ranking: C.

Table 1. Research Papers Included in the Dissertation

Research Paper A (chapter 3) draws on the TOE framework (DePietro et al. 1990) and analyzes which factors influence organizations' decisions and ability to adopt AI, as well as how the introduction of AI differs from the introduction of other technologies (see **RO 1**). Twelve interviews with experts focusing on their perceptions of the process of adopting AI were conducted in accordance with a qualitative research approach. The results show that the TOE framework should be expanded to include additional factors that emerged in the analysis of the interviews. Among these factors are data-related aspects as well as regulatory issues (e.g., the General Data Protection Regulation). Overall, the results of our study provide a basis for future AI adoption research and can serve as guidance for managerial decision-making regarding the introduction of AI.

Research Paper B (chapter 4) uses qualitative interviews and data scraped from the data science community Kaggle to explore why organizations host data science competitions and when organizations perceive them as successful (see **RO 2**). The results show some benefits related to data science competitions, such as learning new approaches and technical discussions among participants; however, conducting such competitions is very time consuming, and outcomes are neither certain nor integrated into a holistic solution. In total, 12 factors that influence an organization's perceptions of the success of hosting data science competitions were identified.

Research Paper C (chapter 5) draws on the theory of innovation diffusion (Rogers 2003) to investigate whether users perceive AI-based advisory systems as providing a relative advantage over human advisors and which mechanisms establish trust in AI-based advisory systems (see **RO 3**). Using an online survey in the context of financial planning, we collected data from 226 participants. The results show that users appreciate the convenience that AI-based advisory systems offer by providing easy and instant satisfaction of informational needs. Furthermore, we found that 11 measures to increase trust in AI-based advisory systems could be classified into three categories based on their effectiveness. Non-committal testing improved trust the most, while the implementation of human traits was least effective.

Research Paper D (chapter 6) is based on task–technology fit (TTF) and JAS and analyzes users' actual advice-taking behavior in the context of financial robo-advisors (see **RO 4**). We conducted an experimental study among 197 participants measuring actual advice-taking behavior and analyzed the data using group comparisons and structural equation modeling. Our results show that the perceived advisor's expertise is the most influential factor in the task–advisor fit for robo-advisors and human advisors. Furthermore, the perceived advisor's integrity (only for human advisors) and perceived efficiency-enhancing capabilities (only for robo-advisors) affect the task–advisor fit. Additionally, we found that task–advisor fit affects users' actual advice utilization.

Research paper E (chapter 7) uses the design science method to develop an IT artifact that supports researchers in conducting structured literature reviews using ML (see **RO 5**). The artifact aims to partially automate the literature review process, from collecting documents through analysis. Documents are downloaded from different databases based on defined keywords, processed using optical character recognition and the word2vec algorithm, and analyzed with latent Dirichlet allocation (LDA) topic modeling, rapid automatic keyword extraction, and agglomerative hierarchical clustering. Finally, illustrations such as dendrograms are used to visually represent the results.

2 Theoretical Fundamentals

This chapter provides an overview of the theoretical foundations of this dissertation, whereas the theoretical background related to each paper is presented in more detail in chapters 3 through 7.

2.1 Artificial Intelligence and Machine Learning

AI refers to “the science and engineering of making intelligent machines, especially intelligent computer programs” wherein “intelligence is the computational part of the ability to achieve goals in the world” (McCarthy 2007, p. 2). A more tangible definition of AI describes it as the study of intelligent agents that can perceive their environment and perform actions (Russel and Norvig 2009). How an intelligent agent maps perceptions and actions can be determined by a variety of functions (e.g., rule-based, model-based) (Russel and Norvig 2009). A common approach to generating a mapping function that is often associated with AI is ML. ML can be defined as an approach to deriving patterns from data using algorithms to create a model of reality (Mitchell 1997). Ideally, these models can be used to determine an agent’s best possible action given its perceptions: the so-called “rational action” (Russel and Norvig 2009).

A common definition of ML is provided by Mitchell (1997, p. 2), who states, “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.” ML approaches can broadly be categorized into three classes: supervised learning, unsupervised learning, and reinforcement learning (e.g., Bishop 2006; Jordan and Mitchell 2015). Supervised learning uses training data that contains examples of corresponding input and target vectors (i.e., labeled data) (Bishop 2006). In classification problems, the output vector consists of a finite number of discrete categories, whereas in regression problems, the output vector contains continuous variables (Bishop 2006). Unsupervised learning is characterized by training data in which the input vectors do not have corresponding target vectors (Bishop 2006). It can be applied to (for example) clustering (i.e., discovering groups of observations that are similar) or anomaly detection (i.e., identifying observations that deviate significantly from the majority of the data) (Bishop 2006). Reinforcement learning is a technique wherein an agent learns from a series of reinforcements (i.e., rewards or punishments) (Russel and Norvig 2009). Often, this is operationalized by an agent that interacts with an environment and evaluates the outcomes of its actions (Bishop 2006). Put simply, the agent learns by trial and error. Exemplary applications of reinforcement learning include agents learning to play games such as Go or Dota 2 (Berner et al. 2019; Silver et al. 2016) or robotics (e.g., Kober et al. 2013).

Recent research on AI focuses on the development of artificial neural networks (ANN) with increasingly complex architectures—referred to as deep learning—and their application (e.g., Girshick et al. 2014; Goodfellow et al. 2014; Krizhevsky et al. 2012). In general, ANNs are composed of artificial neurons that are typically arranged in multiple layers. Each neuron has inputs and calculates one or multiple outputs. Typically, each layer contains multiple neurons and ANNs consists of at least three layers—an input layer, one or multiple hidden layer, and an output layer (Goodfellow et al. 2016; LeCun et al. 2015). This multi-layered architecture allows them to automatically discover representations from raw data (e.g., images, speech) and process it for detection or classification (LeCun et al. 2015). Novel deep learning approaches are responsible for most of the recent breakthroughs in ML such as image generation, machine translation and computer vision (e.g., Gregor et al. 2015; Johnson et al. 2017; Redmon et al. 2016). However, the high performance of deep learning comes at certain costs. For one, training such networks takes a lot of time, money and energy, resulting in a massive carbon footprint (Schwartz et al. 2020). Furthermore, due to their nested and non-linear structure, ANNs lack of transparency, which is why they are referred to as “black-boxes” (Samek et al. 2017). However, rapid progress is made in the field of explainable AI, which seeks to develop techniques that enable humans to understand such black-box models (Guidotti et al. 2018; Lundberg and Lee 2017; Ribeiro et al. 2016).

2.2 Use of Information Systems

In IS research, the acceptance, use, and adoption of technology constitute a major research stream around one of the most central and most widely studied constructs (Burton-Jones et al. 2017; Córdoba et al. 2012; Straub and del Giudice 2012). The use of IS can be defined as an actor’s employment of an IS to perform a task (Burton-Jones and Gallivan 2007). Under this definition, an actor refers to an individual, group (or any other collective), or even entire organization (Burton-Jones et al. 2017). Originating from the application and integration of many social psychological findings, such as the theory of reasoned action (Fishbein and Ajzen 1975), the research on IT acceptance and use developed robust theories such as the technology acceptance model (TAM) (Davis 1989) and the unified theory of acceptance and use of technology (UTAUT) (Burton-Jones et al. 2017; Venkatesh et al. 2003, 2012). From this point, research evolved and gradually refined the understanding of IS use by focusing on each of its elements (e.g., actors, tasks, technology) (e.g., Bhattacherjee and Premkumar 2004; Rai et al. 2012; Sarker and Valacich 2010). Two of the most important constructs that these theories have in common relate to the benefits and efforts associated with the use of the technology. This emphasizes the relevance of design science to creating IS that can be used effectively and efficiently (Hevner et al. 2004; Tsichritzis 1997). Therefore, a complementary research cycle between behavioral science and design science can help address fundamental problems in the productive application of IS (Hevner et al. 2004).

The TTF model is a widely accepted approach to investigating IS utilization at an individual level. It is a well-studied model that seeks to understand the mechanisms of IS utilization and its impact on task performance (Goodhue 1995; Goodhue and Thompson 1995a). TTF refers to the state wherein a technology provides features and support that “fit” the requirements of a specific task and is affected by technology and task characteristics (Goodhue and Thompson 1995b). It has already been widely applied in various contexts to investigate the success of new technologies such as online shopping, question-answering machines, and management information systems (e.g., Goodhue et al. 2000; Klopping and Mckinney 2004; Robles-Flores and Roussinov 2012).

As a complement, the TOE framework can be used to investigate the adoption and implementation of IS at an organizational level. As a framework, it provides a useful and flexible starting point to study the adoption of innovations (Zhu and Kraemer 2005). In essence, the framework suggests that the process by which a firm adopts and implements innovations is affected by its technological, organizational, and environmental context (DePietro et al. 1990). The technological context refers to internal and external relevant technologies, the organizational context describes internal structures and processes, and the environmental context refers to the external regulations and conditions imposed on the organization (DePietro et al. 1990). The TOE framework has successfully been applied to other contexts, such as business intelligence systems (Hatta et al. 2017) and big data (Bremser 2018).

2.3 Research Methods

This section provides an overview and discussion of the research methods selected for the five research papers included in this dissertation (chapters 3–7).

A common and widely accepted categorization for research methods is the distinction between qualitative and quantitative methods (Myers 1997). Qualitative research methods stem from the social sciences, where they are used to study social and cultural phenomena by using non-standardized data such as interviews, documents, or observations (Myers 1997). As an open and flexible approach, qualitative research is especially useful for collecting more concrete and plastic insights into processes and interrelationships; it is therefore well suited to exploring complex topics that have not yet been extensively investigated (Flick et al. 2004). Common qualitative research approaches include case studies, action research, design science, and grounded theory studies (e.g., Corbin and Strauss 2014; Yin 2009).

Quantitative research methods were developed in the natural sciences and rely on the analysis of numerical data (Myers 1997). As a structured and standardized approach, quantitative research aims to derive general relationships using statistical analysis. It especially provides the benefits of a controlled experimental setting, careful measurement, and generalizable samples (Miles and Huberman 1994). Common quantitative research methods include surveys, numerical methods, laboratory experiments, and formal approaches (Myers 1997).

Both types of research are widely accepted in the IS discipline and are applied in the studies included in this dissertation. Figure 2 provides an overview of the various research designs used in the studies.

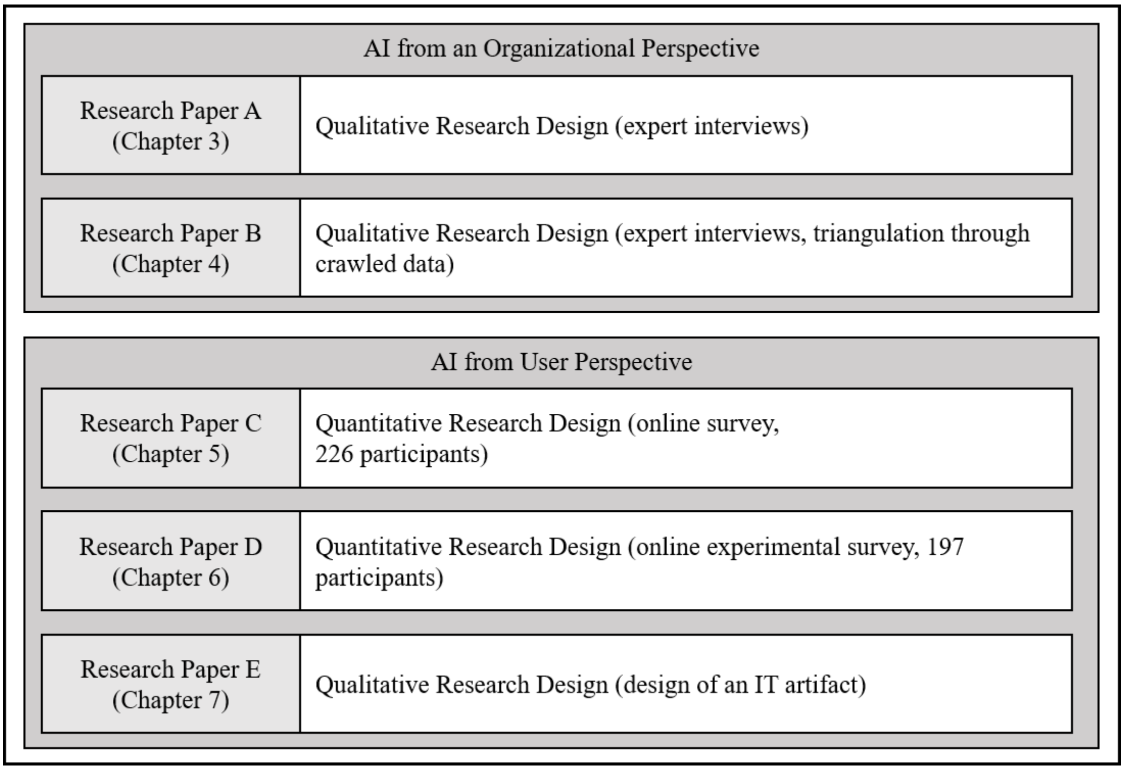

Figure 2. Overview of Research Design and Methods Used in the Dissertation

To analyze the readiness factors of organizational AI adoption and implementation (Research Paper A), a qualitative research approach using expert interviews was selected due to the exploratory nature of the study. Using the TOE framework and factors from extant literature as a conceptual starting point, the interviews were used to confirm and expand the initial conceptualization and to obtain complementary views on the topic.

Qualitative expert interviews were also used to investigate organizations' motivations to host data science competitions (Research Paper B). This approach was deemed appropriate because the amount of organizations hosting such challenges is rather small and the topic has not been extensively explored. The interview data was enriched with publicly available data crawled from the data science platform Kaggle, where most of these competitions are hosted. A more holistic view on the topic was thus gained, and interview statements could be validated.

In Research Papers C (chapter 5) and D (chapter 6), a quantitative approach was selected. Specifically, online surveys were chosen for data collection because this research method allows for a standardized, controlled environment, careful measurement, and results that are generalizable to a population of interest.

Finally, Research Paper E (chapter 7) employs a qualitative method by using design science research to develop an IT artifact that uses ML to support the process of structured literature reviews. The artifact was developed in an iterative manner by continuously assessing requirements, implementing features, and evaluating the results.

3 Paper A: Exploring Organizational Readiness Factors for Artificial Intelligence

Title

A New Organizational Chassis for Artificial Intelligence - Exploring Organizational Readiness Factors

Authors

Pumplun, Luisa, Technische Universität Darmstadt, Germany

Tauchert, Christoph, Technische Universität Darmstadt, Germany

Heidt, Margareta, Technische Universität Darmstadt, Germany

Publication Outlet

Proceedings of the 27th European Conference on Information Systems (ECIS 2019), Stockholm, Sweden

Abstract

In 2018, investments in AI rapidly increased by over 50 percent compared to the previous year and reached 19.1 billion USD. However, little is known about the necessary AI-specific requirements or readiness factors to ensure a successful organizational implementation of this technological innovation. Additionally, extant IS research has largely overlooked the possible strategic impact on processes, structures, and management of AI investments. Drawing on TOE framework, different factors are identified and then validated conducting 12 expert interviews with 14 interviewees regarding their applicability on the adoption process of artificial intelligence. The results strongly suggest that the general TOE framework, which has been applied to other technologies such as cloud computing, needs to be revisited and extended to be used in this specific context. Exemplary, new factors emerged which include data – in particular, availability, quality and protection of data – as well as regulatory issues arising from the newly introduced GDPR. Our study thus provides an expanded TOE framework adapted to the specific requirements of artificial intelligence adoption as well as 12 propositions regarding

the particular effects of the suggested factors, which could serve as a basis for future AI adoption research and guide managerial decision-making.

Keywords

artificial intelligence, adoption, TOE framework, organizational readiness

3.1 Introduction

"The world's most valuable resource is no longer oil, but data" – proclaimed by The Economist (2017) and a plethora of other articles, the business value of data is widely accepted. If data is the new oil of our economy and artificial intelligence (AI) is fuelled by data, then AI can analogously be referred to as the engine (Agrawal et al. 2018). Thanks to improved algorithms in deep learning and ample access to historical datasets as well as cost-effective computing power and storage space, AI applications are on the rise and receive increasing attention from both technology companies and more 'traditional' companies that anticipate competitive advantages (MSV 2018). Despite inconspicuous short term impact, long term commitment is important since AI represents a paradigm shift for organizations (Hosanagar and Saxena 2017). According to Gartner, "85 percent of CIOs will be piloting AI programs through a combination of buy, build, and outsource efforts" by 2020 (Andrews et al. 2017, p. 2) – however, just like a new engine for electric vehicles requires a new chassis, approaching an organizational AI project requires an assessment whether the focal organization possesses the necessary prerequisites and framework to enable successful AI initiatives.

Despite ever increasing organizational (and governmental) investments in AI (Bughin et al. 2017), less than 39 percent of all companies have an AI strategy in place, only 20 percent of companies have actually incorporated AI in some offerings or processes, and merely 5 percent have extensively incorporated AI (Ransbotham et al. 2017). The easiest explanation for this apparent hesitance are prominent examples of AI projects gone awry, like the Microsoft Chatbot Tay tweeting racist slurs (Reese 2016) or IBM's Watson failing to diagnose cancer as promised in their advertising campaign (Flam 2018). However, most so-called AI failures cannot be attributed to AI itself but rather to the underlying processes and the involved people. Current AI research has focused predominantly on technical advancements (e.g., Lu et al. 2018; Monroe 2018) but largely factored out the necessity to analyse the readiness of the 'organizational chassis' to successfully support AI initiatives. In this regard, AI initiatives cannot be approached like yet another new technology trend since several aspects distinguish these projects from previous technology initiatives, e.g., cloud computing adoption or social media marketing: in its essence, AI refers to a broad and complex set of approaches that do not have to confine themselves to methods that are observable and have thus been often compared to a black box (McCarthy 2007). In accordance with McCarthy (2007, p. 2) we understand AI as a "science and engineering of making intelligent machines, especially intelligent computer programs", which tries but is not limited to simulate human intelligence and which includes underlying technologies like machine learning, deep learning and natural language processing

(Elliot and Andrews 2017). AI differs from non-AI technology as it learns to make decision based on incoming data, rather than being based on an explicitly defined set of rules (Crowston and Bolici 2019). This self-adaptive property allows AI to learn from user behaviour, react to its environment, and make complex decisions automatically. These properties result in human attributes being assigned to AI (Rzepka and Berger 2018). However, the technology is also perceived as a threat because the algorithm's decision is not transparent (i.e., black box behaviour) and is likely exceeding human capabilities in a particular task due to its efficiency and scalability (Brundage et al. 2018).

In an information systems (IS) context, researchers have only recently begun to examine organizational readiness factors for AI (e.g., Alsheibani et al. 2018) but have as of now not yet expanded frameworks like TOE (technological-organizational-environmental) to cover the specific characteristics AI initiatives entail across industries and adoption stages. Due to the scarce extant literature, this study explores organizational readiness factors through a qualitative interview approach with 14 experts from both user and provider firms at various adoption stages. Building on TOE as conceptual framework, our approach thus aims to identify:

(1) Which factors influence the decision and the ability to adopt AI in organizations? And sets out to shed further light onto (2) What explicitly distinguishes the introduction of AI from other technologies?

The remainder of this manuscript is structured as follows: To begin with, we provide a brief overview of the related work and theoretical background (TOE) to mark off the research area before the qualitative study design is presented. After introducing our study sample comprising 14 interviewees, we derive empirical results which are integrated to expand the TOE framework. The results of our paper are a first step in providing a holistic view of the factors that are relevant for adoption of AI in the nascent research landscape. Thereby, the discussion of our key findings illustrates contributions to research and practice and an approach to future work. Finally, we conclude the manuscript by pointing out the limitations of our study and providing specific avenues for future research.

3.2 Theoretical and Conceptual Background

3.2.1 Artificial Intelligence and Adoption

The nascent ubiquitous adoption of AI in companies is currently omnipresent in research and practice, which indicates the potential attributed to AI. However, only few studies have dealt with the organizational aspects of AI adoption like the implementation of the technology into organizational processes and governance structures (e.g., Ransbotham et al. 2017). Extant published studies rather focus on the improvement of this technology and its underlying algorithms (e.g., Monroe 2018; Yan et al. 2016) or the impact of AI on specific industries and departments (e.g., Huang and Rust 2018; Kruse et al. 2019; Moncrief 2017) – whereas overarching aspects like the influence on AI applications exerted by an organization's strategy or the macro-environment, have scarcely been taken into account in information systems (IS) literature (Nascimento et al. 2018).

Indeed, a literature review by Nascimento et al. (2018) demonstrates possible avenues for future studies by identifying specific aspects which should be considered when adopting AI technologies (i.e., high commitment to the area, human requirements to deal with the techniques), but they do not integrate their findings into a theoretical framework. Similarly, Rzepka and Berger (2018) focus on the interaction of AI systems and users and address important factors (e.g., the fit between the user, system and task), but do not apply a distinct adoption framework. There are some further, rather practice-oriented contributions analysing or discussing the adoption of AI. For example, vom Brocke et al. (2018) state that new job profiles have to be created, resulting in the necessity of adequate skill development of employees and the adjustment of corporate strategies.

However, the aforementioned findings are still rather disparate and do not provide a concise framework that could guide future organizational studies regarding AI and the actual implementation of AI in companies. To the best of our knowledge, there are only two contributions that consider the adoption of AI in organizations from a more theoretical perspective and across various industries (Alsheibani et al. 2018; Rana et al. 2014). Alsheibani et al. (2018), a research-in-progress publication, draw on the TOE framework (DePietro et al. 1990) to explain an organization's readiness to introduce AI into their organization. In line with the existing theory, they constitute technological (T), organizational (O), and environmental (E) factors, which influence AI adoption and propose a quantitative, thus confirmative, approach. Accordingly, influencing factors are selected on the basis of assumptions from past studies, which are not specified in more detail, and on the basis of previous technologies, which do not have the same specific characteristics as AI. Rana et al. (2014), on the other hand, use the Technology Acceptance Model (TAM) to explain the organizational adoption of machine learning techniques in the specific context of software defect prediction. Again, the unique characteristics of AI are not sufficiently addressed. Instead, existing concepts (e.g., perceived benefits) are examined based on a sample of only four interviewees from two companies. Given that AI differs from previous technologies in several ways, an all-embracing framework needs to take these differences into account (Zhu and Kraemer 2005): AI is considered both efficient and scalable, is able to exceed human capabilities and comprehension (Brundage et al. 2018), derives its own rules from added data (Crowston and Bolici 2019) and shows a distinctive black box behaviour (Adadi and Berrada 2018). In addition, recent developments affect the organizational use of AI (e.g., improvement of deep learning algorithms) making it necessary to collect comprehensive, up-to-date data.

Since no current exploratory study investigates the adoption of AI across various industries, an explorative approach is necessary to provide further insights that potentially deepen and extend the proposed TOE framework to account for the novelty regarding the organizational implementation and adoption of AI.

3.2.2 TOE Framework and Diffusion of Innovation

In general, the TOE framework represents a useful and somewhat flexible starting point to study innovations as it provides a generic theory for the diffusion of technologies (Zhu and Kraemer 2005). Therefore, it has been widely applied to other contexts and technologies like cloud computing (e.g., Lian et al. 2014), big data (e.g., Bremser 2018) and business intelligence systems (e.g., Hatta et al. 2017). In essence, the TOE framework comprises three main elements that influence the adoption process of technological innovations: (a) the technological context describing the internal and external relevant technologies available, (b) the organizational context that depends on internal structures and processes measured by various factors such as company size and free resources and (c) the environmental context, which describes the business related field of action, taking into account industry, competitors, government, and suppliers (DePietro et al. 1990). Following Zhu and Kraemer (2005) the TOE framework can be extended by using the innovation diffusion theory of Rogers (1995), which states different technological factors including relative advantage and compatibility. Relative advantage is described as the degree to which an organization perceives an innovation better compared to the previous solution. The second factor, compatibility, is the degree to which an innovation matches the actual needs of the potential user organization. Both factors are positively related to its rate of adoption (Rogers 1995). Looking at the organizational readiness, DePietro et al. (1990) postulate a positive influence of the strategic behaviour of management, organization's size and slack resources. They also point out the relevance of the intensity of competition as a positive factor on adoption as well as governmental regulations, which can have both, negative and positive effects on innovation implementation. Since there is only little research on AI adoption, a general TOE framework as described above is used as an initial conceptual starting point (see Figure 3), which will be expanded in the course of the study.

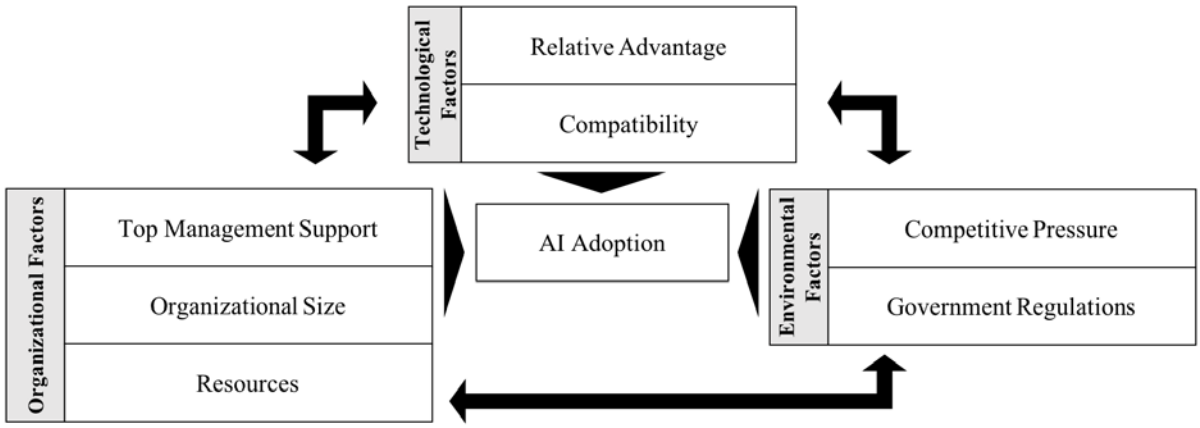

Figure 3. TOE Framework as Conceptual Base (based on DePietro et al., 1990; Rogers, 1995)

3.3 Qualitative Research Methodology

The aim of the study is to expand the current state of IS research concerning AI application in organizations by questioning experts who work on managerial and operational levels for AI provider and user firms. Organizational AI adoption is a complex topic and has not yet been fully explored. Therefore, an explorative approach using interviews with experts seems appropriate to investigate the problems occurring in this particular context (Flick et al. 2004). According to Weber (1990), content analysis can be used to assess open-ended questions, making the approach suitable for evaluation of the collected qualitative data. Thus, in order to develop an organizational adoption framework, this paper follows the steps of content analysis (see Figure 4): Based on the TOE framework, which serves as a conceptual framework, seven initial categories were derived from relevant literature (e.g., factors "compatibility" or "top management support" in Figure 3). By analysing the interviews, these categories are examined and extended gradually, resulting in 23 categories and subcategories of the final framework for AI adoption. The interviews are transcribed, coded and analysed taking into account relevant practice-oriented studies through triangulation (Hsieh and Shannon 2005). In particular, we use a combination of directed and conventional analysis, where the directed approach uses codes derived from theory (i.e., TOE framework) and the conventional analysis takes into account information obtained directly from the data since the applied theory is not specifically adjusted to AI technology and therefore should be supplemented and deepened inductively (Hsieh and Shannon 2005).

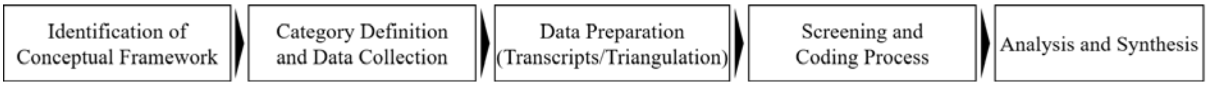

Figure 4. Content Analysis Process (based on Hsieh and Shannon, 2005)

3.3.1 Research Design

Our main information source were in-depth expert interviews, which were conducted in a semi-structured way. Thereby, the guiding principles of Sarker et al. (2013) were considered by preparing an interview protocol and questioning key informants in different companies. In order to avoid typical pitfalls of semi-structured qualitative interviews, contact was established with the interview partners via e-mail and telephone before the interviews were carried out (Hermanns 2004). While conducting the interviews we kept our questions open in order to enable participants to speak freely.

The interview guide comprises three different sections. The first section comprised general questions about the position and responsibility of the interviewee and their previous experience in the field of AI and related technologies in an operational or managerial context. The second and most comprehensive section considered advantages and risks of using AI (i.e., the possible results of AI initiatives) and the triggers, prerequisites and limitations of using this technology in organizations. In addition, we inquired the criteria used by the companies to assess the general potential of AI. The last set of questions dealt with the actual use of AI and the strategic and tactical challenges it poses. For example, we asked the interview partners which AI-based applications are currently being used and which specific actions were associated with the introduction and implementation of these projects. Due to the semi-structured approach, initial questions were subject to a gradual adjustment in order to account for the individual expertise and position of the participants and to develop the focus during the interviewing process.

3.3.2 Sample and Data Collection

We provide an overview of the participants in Table 2 (see below) and further details in the following.

Participants (UF): Participants of firms that are predominantly users of AI products and services					**Participants (PF):** Participants of firms that are predominantly providers of AI products and services				
ID	**Position**	**Job Exp.**	**Interview Method**	**Adoption Stage**	**ID**	**Position**	**Job Exp.**	**Interview Method**	**Core/ Non-core**
P-01	Digital Growth Manager	16 years	Face-to-face	Adoption	P-08	Founder	10 years	Face-to-face	C
P-02	Head of Marketing & Analytics	10 years	Face-to-face	Consider-ation	P-09	Development Manager	6 years	Face-to-face	C
P-03	Head of Dig-ital Commu-nications	14 years			P-10	Solution Manager	15 years		
P-04	Asset Management Strategist	3 years	Telephone/ Face-to-face	Adoption	P-11	Development Manager	7 years	Face-to-face	C
P-05	Chief Product Owner	8 years	Face-to-face	Continued use	P-12	Managing Director	19 years	Written answer	NC
P-06	Product Owner	8 years	Face-to-face	Continued use	P-13	Consultant	2 years	Telephone	C
P-07	Account Executive	3 years	Telephone	Adoption	P-14	Managing Director	11 years	Telephone	C
Awareness: Org. becomes aware of AI **Consideration:** Org. considers to adopt AI **Intention:** Org. intends to adopt AI **Adoption:** Org. begins to adopt AI **Continued use:** Org. continues to use AI					**Core (C):** AI capabilities and products differentiate company strategically from others **Non-Core (NC):** AI capabilities and products are no strategic factor for company				

Table 2. Participant Overview

The interview partners were selected on the basis of a key informant approach. Following the rules of data triangulation, both user (UF) and provider firms (PF) were surveyed (Flick 2004). The answers were collected over a six-month period and took place between May and October 2018. In total 12 interviews with 14 highly involved participants were conducted within two European countries (Germany and Ireland), taking into account seven experts from provider firms and seven experts of companies, which mainly purchase AI products. After the 12th interview, data collection was discontinued as a further contribution of additional qualitative data was considered unlikely (i.e., theoretical saturation was assumed) (Flick 2004).

Among the 14 interviewees were eleven male and three female participants. The total number of respondents is comparable to other qualitative studies that consider the adoption of similar technologies (e.g., Bremser et al., 2017; Mallmann and Gastaud Maçada, 2018). In order to avoid an elite bias, both IT staff and managers were interviewed (Miles and Huberman 1994). Therefore, three of the participants were managing directors or founders, eight identified as middle managers or heads of departments, while the remaining respondents were either consultants or strategists. For the purpose of potentially achieving more generalizable research results and identifying sector and enterprise size-specific differences (Flick 2004), companies across various industries and of differing sizes were selected, including large (75 %), medium-sized (17 %) and very small enterprises (8 %) (European Commission, 2003) from industries like electricity, gas, steam and air conditioning supply (D), information and communication (J), manufacturing (C) as well as wholesale and retail trade (G) (United Nations 2008). At the time of the interviews, the organizations were in different phases of implementation regarding AI. Based on the classification according to Frambach and Schillewaert (2002), user firms are divided into the following stages of adoption: awareness, consideration, intention, adoption and continued use, while provider firms were classified according whether they offered AI as a core competence or not (Leonard-Barton 1992).

The interviews lasted on average 58 minutes and were mainly held face-to-face because of the complexity, scope, and sensitivity of the topic. Nevertheless, a total of four interviews were conducted using telephone calls and one participant replied in a written form due to geographical distance. An overview of the surveyed participants can be found in the table above (see Table 2).

3.3.3 Coding Concept

Most of the interviews were recorded and transcribed after agreement by the interviewees. In a single interview only notes were taken and in another case a written answer was submitted. Subsequently, the transcripts were assessed by using the NVivo 12 software and by conducting two coding cycles as recommended in Saldaña (2009). The first coding cycle comprised a mixture of attribute coding, descriptive coding and hypothesis coding. The former is performed to obtain essential insights about the data and its descriptive information (e.g., UF/PF, size of organizations). In addition, hypothesis coding was carried out to account for the initially conceptualized factors from the TOE framework (see Figure 3). These factors mentioned in the existing theory form the focus of the hypothesis-based approach and are deductively tested (Greener 2008). Finally, descriptive coding is used to extract additional aspects that go beyond the previously identified factors (e.g., relative advantage, competitive pressure, and top management support) and thus potentially extend the existing framework. In a second cycle, the formerly created codes are combined into a smaller number of sets using pattern coding (Saldaña 2009). By discussing and assessing the coding process with a group of four IS researchers and students, an investigator triangulation helped to ensure rigor and trustworthiness. Furthermore, an ongoing data triangulation process took place while coding the interviews by utilizing multiple sources of evidence (Flick 2004). For example, additional corporate resources as well as current practice-oriented AI studies and reports were considered (e.g., Andrews et al. 2017; Ransbotham et al. 2017).

3.4 Results and Discussion

While validating the proposed TOE framework for adoption of AI (see Figure 3), we found evidence that the established factors do not fully reflect the challenges that companies face when they want to introduce AI to their companies. The presented TOE framework merely includes fundamental factors that are also applicable to other technologies such as cloud computing. Therefore, the findings that do not go beyond these basics are summarized in tabular form (Table 3). Aspects that supplement or contextualize the original framework will be examined in more detail below.

El.	Fact.	Results	Statements
Technological Factors	Relative Advantage	With the help of AI it is possible to learn from the data over time. However, AI is not a panacea, but should be compared to the use of robust conventional systems for the specific application. The combination of both approaches should also be considered in order to solve the overall problem. This assumption is strengthened by Rzepka and Berger (2018), who indicate that AI is better suited for particular use cases than others. In addition, it is demanded that the results of AI be made comprehensible and no longer represent a black box. The demand for more transparency of AI based systems is also demanded in the current IS literature (e.g., Crowston and Bolici 2019; Rzepka and Berger 2018).	"But that one adapts, that one learns based on collective knowledge, no matter if one provides it now at the beginning or continuously, that one adapts there then, that is actually the strength of this AI." – P-11 "And it may well be that you get on with workflows or get on with fixed processes. Or that you say, you know what, we just run AI in the background. And we just take a look at which needles the system still brings us. But it's by no means a panacea […]." – P-01 "We know that we can't really understand machine learning. […] And that there must be procedures that show that exactly this one feature was responsible for it." – P-13
	Compatibility	For the successful use of AI, the work processes must be adapted to the technological requirements. Furthermore, there must be a fit between the desired application and technology.[a]	"If I then ask [...] why do the projects fail? You then realize that the need was not clearly communicated, the use case was not right, that it was too big. That you say you want to do something, but you don't know what." – P-01
Organizational Factors	Top Management Support	In principle, the support of top management can facilitate the introduction of AI. However, a certain understanding of the technology and its applications is required. Currently, decision makers in middle management are particularly problematic, as they are very KPI-driven and thus inhibit AI use.	"Someone, a top manager or someone comes from some conference, has picked up something like Big Data or Predictive Maintenance as buzzwords and then says, 'yes, let's do it'. Yes? And then you start to code somehow and you start to collect and somehow you notice then hey, actually we don't know exactly what we are supposed to do now." – P-14
	Organizational Size	It is unclear whether larger companies have a better chance of adopting AI. Basically, a high budget and a large volume of customer data enables and justifies the use of AI. However, the slow group structures are also hampering further development in this area.	"Now are you going to [...] I'd rather say a niche area. Niche in the sense of, you have maybe only 10,000 users. Then it's not worth the effort that data scientists, Computational Linguists develop something for five years." – P-11

	Resources	The resources can be divided into the factors budget, employees and data that affect the use of AI.[a]	“I think obstacles […] are certainly the initial expenditures. At the beginning, you’d need a small one-off budget, a bit of know-how as a starting point […].” – P-02
Environmental Factors	Comp. Pressure	Competitive pressure leads companies to increasingly deal with AI in order to gain a competitive advantage.	"They [the costumers] challenge us too. They say, look at the competition, the start-up does that, we've already looked with them. Why can't you do that yet?" – P-10
	Gov. Regulations	Many laws complicate the introduction and use of AI. In this context a renewal of the legal situation is demanded. Especially the GDPR and the employees’ council are a particular hurdle for companies.[a]	“And innovation and law are two words that I think rarely appear in one single sentence.” – P-04
[a]: Further details on the subcategories are discussed below			

Table 3. Findings: Examination of Proposed Factors in TOE Framework

In addition to the ‘classic’ TOE assumptions, the experts also mention prerequisites for the implementation of AI that result from the special properties of AI and therefore have only been insufficiently addressed or have not been examined in general TOE literature at all before. These new findings are described comprehensively in the following section.

Technological Factors

Technological factors comprise two main aspects: Relative advantage, which was already considered in Table 3 in detail, and compatibility, which can be divided into two subcategories on the basis of expert interviews: business processes and business cases. Therefore, we will revisit the second factor compatibility in the following and explain it in more detail.

Compatibility. According to experts, the *business processes* in the company must be adapted to the new requirements that arise from the use of AI. In the context of AI, it is therefore no longer useful to use existing KPIs of other projects, since AI projects have differing properties. For example, the results that arise from such projects can no longer be planned to an extent that would be necessary regarding traditional, common KPIs (e.g., ROI) as demonstrated by the following quote:

"The interesting thing about how we implement these projects here is that we didn't define KPIs […]. That means for us, we learn with the information we get back through the system. That's a very important point. If you apply old KPIs to new technologies and approaches, you run the risk of only digitizing old KPIs." – P-01

Instead, it becomes necessary to introduce agile forms of work. Particularly in the field of data science, it is important to continuously evaluate the progress of projects, since the feasibility of ideas in this area cannot be proven from the outset. There are only a few, incomplete criteria to evaluate the existing data at the very beginning. Within the framework of agile, flexible working models for software development, the current status and the data can always be viewed in terms of new findings, thus reducing the risk of investing the wrong amount of time and money. The relevance of agile working methods is underlined by the following statement:

“And in IT you had very, rigid waterfalls, that is classic traditional IT project management. Which is not, how shall I say, very beneficial regarding the uncertainties when using data and artificial intelligence. […] Because you just plan a concept somehow, that's actually this classic process, over half a year and then you look into the data and notice ‘oh God, that's all wrong!’. And you can actually throw the concept away! So half a year, more or less, not as much progress has been made as if one had looked at the data in advance." – P-14

In addition to the work processes, however, further factors must also be checked for compatibility. Another very frequently mentioned aspect is the formulation of a concrete *business case*. Experts believe that AI can only be used successfully if there is a clear problem. AI must be seen as a tool for a purpose and cannot be viewed in isolation. The problem of prioritizing possible use cases appropriately is known from literature on big data use (e.g., Bremser 2018), which also deals with an underlying technology that can be used in a variety of ways in organizations.

“But you really need to know, ‘where can you solve a problem with that?’. Just because you can do AI, it doesn't bring you anything, zero, honestly not. […] They don't buy it because it's AI. So really, also corporate customers, they don't buy it because there is AI in it now. They buy it because it must have a benefit.” – P-08

In line with these factors influencing AI adoption, we formulate the following propositions:

Proposition 1*: Compatibility between AI technology and business processes (e.g., agile forms of work) as well as the development of a dedicated business case will have a positive effect on adoption of AI in companies*

Organizational Factors

In addition to technological readiness, factors must also be taken into account that reflect the overall organization's ability to implement AI. The factors culture and organizational structure were newly discovered by examining the expert interviews, while the factor resources was subdivided into the aspects budget, employees, and data.

Culture. After evaluating the interviews, it became evident that the adoption of AI in a company is strongly influenced by the culture in the company. In addition to top management support the introduction and implementation of an innovative culture in the company are also relevant.

In this context, aspects of *change management* to achieve an innovative culture within the company were mentioned frequently by the interviewees. The functionality of an intelligent application is based on the input of already existing, high-quality data as well as the training which has to be carried out by the employees over time (Crowston and Bolici 2019). Only if there is a willingness to use the technology in the long run, the quality of the answers and decisions made by the machine will improve.

"In the beginning the model is bad. You have few answers that reach this threshold. But by constantly saying as an employee that this was right or by correcting, you are building a knowledge base." – P-08

If the path to an *innovative culture* is not successful, there is a danger of missing out on new, important technologies and trends. The factor of missing an absorptive capacity to adopt new technologies is evidenced by the following statement:

"In such a large corporation you have the tendency to say again and again 'well, we make money with the model we have! Why should I come up with something new now?'." – P-05

Resources. The adoption of AI in a company does not only depend on the culture, but also results from slack resources, which should be further subdivided. Comparable to other innovations (Bremser et al. 2017), the available financial resources through a *budget* are an important aspect that generally determines the implementation of new technologies in projects. A high budget can enable capacities, create financial freedom and help to build know-how. On the other hand, obligations also arise from financial resources. This problem can in turn jeopardize the successful introduction of AI, since the course of projects with AI is unpredictable and strongly dependent on the data used. The restricting influence of budget is demonstrated by the following statement:

"The second point is the budget. The moment your management or the person responsible for the budget asks the question 'what is the return on investment?'. And 'what happens if I don't do it?' You are no longer on the move agilely, but you are immediately arrested in a major project. The demand or the requirements are already defined, there's a price tag on it and there's a timeline on it. No more room for adjustments." – P-01

In addition to the budget, a second aspect should be considered as one of the most frequently discussed factors within the sample: the *employees* of a company who have the necessary know-how to apply the technology. Basically, it should be noted that the staff should have both, the professional qualifications and programming knowledge in the field of AI (e.g., utilizing libraries such as TensorFlow, PyTorch or Keras) as well as a domain understanding of the respective organization. It should also be considered that many companies have problems recruiting professionals such as data scientists, who demand high salaries and are potentially disloyal to their employers due to a high demand on the labour market. The necessity of these occupational groups for implementation of AI is also addressed by previous studies (e.g., Kruse et al. 2019). Additionally, interviews show that AI projects cannot simply be outsourced as they require the company's domain knowledge as described. Therefore, an expert suggests to train the employees in the company who already have a domain specific knowledge (e.g., controller, statisticians) in the field of machine learning. The problem set is evidenced by the following statement:

"This is one of the most important things: you need the people! In this day and age you can no longer outsource. Especially not with machine learning and artificial intelligence. That doesn't work. You need the experts. You need the people – who actually don't have the time." – P-01

The third subcategory that can be seen as a resource is the *data* used to train the AI. Data was among the factors most often mentioned by all interviewees across firms and positions and is also frequently considered in current literature (e.g., Crowston and Bolici 2019). Various problems have been extracted while examining the qualitative interviews: Data must first be made accessible. Both *data availability* and data protection play an important role. Often the data must be made usable from different old systems. Furthermore, it is necessary to extract the data in a scalable form, because AI projects require as many data records as possible. According to the experts, these requirements can account for up to two thirds of the workload of an AI project. The following statement illustrates how time-consuming and difficult the provision of data can be:

"We also often […] first had to think about 'where does the data actually come from?' […] We actually had to deal with three or four different legacy systems from which we had to get the data out." – P-05

In addition to the technical aspects of data availability, *data protection* also plays an important role. Often, it is mainly larger corporations that experience difficulties implementing an open data policy. In these kind of companies, a deliberate isolation of the individual departments takes place, which makes the successful introduction of AI more difficult:

"We're going to have to make sure that we stop pursuing a silo mentality." – P-01

Once the data is available, the quality of the data becomes relevant. This aspect was brought up very often by the interviewees, who point out that *data quality* is regularly a problem, as it is not fully possible to assess the data sets before the project is indeed implemented. Only a few incomplete metrics exist to evaluate the data in advance. This is particularly problematic because historical data often does not have the required quality and degree of detail due to time and cost pressure when data was generated.

“We also have customers who say yes, we have the CRM here, our system here, our old system. Maybe an old application. But we don't really want to take the data with us, because we know that the service staff often just entered something hurriedly due to a lack of time, and that it's not right.” – P-10

Organizational Structure. The culture of the company is closely linked to its structure. As in the statement above, large corporations have problems setting up new AI projects because of their “everything is fine” mentality. Many companies therefore go the way of circumventing old, inhibiting structures by establishing a lab or hub within the organization. However, problems can also arise as a result of this procedure, which is made clear by the following statement by an expert:

“Is this somehow a lab in Silicon Valley, where clever people are all sitting around building something without being subject to the restrictions of the traditional company? The advantage of this is that they are very fast. This has the disadvantage that the integration into the slow company will fail later. [...] On the other hand, if you try it out of the existing IT, which is historically very cost-driven and very innovation-free, then it won't work either.” – P-14

Therefore, it is suggested to use a hybrid model, in which a hub serves as a starting point for new ideas and technologies, but where an intense communication between the lab and the company still exists.

As shown above, organizational readiness factors influence decisions regarding AI adoption of companies strongly. Hence, we posit:

Proposition 2: *A dedicated AI budget, which does not entail any obligations to meet performance targets, will have a positive impact on the adoption of AI in companies*

Proposition 3: *The availability of data scientists and developers with appropriate expertise, domain knowledge as well as the willingness of users to train AI systems over time will have a positive impact on the adoption of AI in companies*

Proposition 4: *The availability of extensive, meaningful and high quality data will have a positive effect on adoption of AI in companies*

Proposition 5: *Departments who keep relevant data to themselves, an overreliance on status quo as well as slow and bureaucratically shaped corporate structures will have a negative effect on the adoption of AI in companies*

Environmental Factors

Looking at environmental readiness, the known factor government regulations is divided into two main aspects (GDPR and employees' council) and the categories industry requirements as well as customer readiness are newly filtered out by coding the expert interviews. The extensions of the original framework are explained in more detail in the following section.

Government Regulations. As already indicated, the introduction of AI must also consider several legal aspects. A relevant regulation that was enforced in May 2018 is the *General Data Protection Regulation (GDPR)*, which regulates activities like the processing of personal data. The handling of the new legal situation is addressed by many experts in the interviews as companies struggle to provide personal data for the training of their intelligent machines. In this context, many data sets need to be anonymized, which makes the use of intelligent, self-learning algorithms more difficult or even impossible. The following statement expresses the impact that such a regulation can have on the European economy:

"This shock with the General Data Protection Regulation […] to make everything bad per se and excessively laborious, that also contradicts any reality. Also, we have to be careful that we don't lose track of others with all these AI topics, because they will do it. We would like to, but we're getting a bit in ourselves' way." – P-11

In addition to legislation concerning the handling of personal customer data, the protection of employees must also be taken into account by firms. Many applications in the field of AI are based on learning from data. If intelligent software is used in the company to support employees, it can access a lot of information from their daily work routine. Thus, there is a danger that the personnel could be monitored. In addition, as a result of the progressive automation by AI, a large scope of duties is taken over gradually by machines. Although it was one of the less prominent constraints mentioned by all interviewees, these effects of intelligent algorithms ultimately lead to the fact that the introduction of AI is inhibited by *employees' council and employee representatives* in companies to protect employees' workplaces.

“Because, of course, a system of this kind, which logs data without limits, could of course also store the information. That X makes three mails in one day and Y makes 30. And her completion rates are much higher. Okay? So the employees’ council is definitely a key stakeholder.” – P-01

Industry requirements. In addition, each industry has its own specific requirements, which also affect the adoption of AI. These are specific laws, external circumstances affecting the company, and the organization's interaction with the environment. For example, Kruse et al. (2019) examine the adoption of AI in financial sector taking into account its specific regulations, IT systems and customer group. These influences can encourage or inhibit the use of AI, depending on their nature. The necessary inclusion of the factor *industry* was evidenced, besides the related literature, by the following statement:

“I also believe that our industry [electricity provision] is simply making a bit of an impact. The challenges facing our industry are simply more complex than what a small retailer might have to solve […].” – P-02

Costumer readiness. When a company is faced with the decision to introduce AI, the knowledge and acceptance of its customer base must also be taken into account. These requirements apply to B2B as well as to B2C companies, which should both focus on their customer’s benefit. The interviewed experts currently see a development of their costumer’s ability and willingness to deal with new technologies. Consumers in particular are increasingly demanding digital and intelligent offers and are acting as disruptors. This is consistent with other adoption literature, which points to changing customer expectations for individualized services and products (e.g., Bremser 2018). But also corporate customers are beginning to innovate. The requirements they will have in the future can be seen from the following statement:

“In 3 to 4 years, when the algorithms are mature, this will become the standard. Then the customers simply expect that such a function [intelligent service] is in the solution.” – P-10

We thus posit that environmental factors, like the legislation or the readiness of industry and customers, affects AI adoption as follows:

Proposition 6*: Strict laws regarding the processing of personal data will hamper the training of intelligent machines and the review by a strong employee representative body will slow down and inhibit the introduction of new technologies. Thereby both will have a negative effect on adoption of AI in companies*

***Proposition 7**: Industry specific properties (e.g., specific regulations, customer group) will, depending on their nature, have both positive and negative effects on the adoption of AI in companies*

***Proposition 8**: Demanding customers will nudge the companies to design individualized, intelligent products and thus will have a positive effect on the adoption of AI in companies*

The previous findings will be used in the following to supplement the basic framework (see Figure 3) and to generate an overview of the experts' statements and thus the special features of AI (see Figure 5).

After the proposed framework has been extensively investigated and extended, the next step is to showcase special features that occur during the introduction of AI in comparison to other technologies and which go beyond the theory of TOE. For this purpose, the statements of experts are investigated via crosstab queries (i.e., filter coded interviews simultaneously by a factor and company type) in order to get an idea about perceptual differences between provider and user firms, which eventually create a gap between supply and demand. The comparison inductively leads to different problem areas where the preconditions, views and attitudes of the provider and user firms differ.

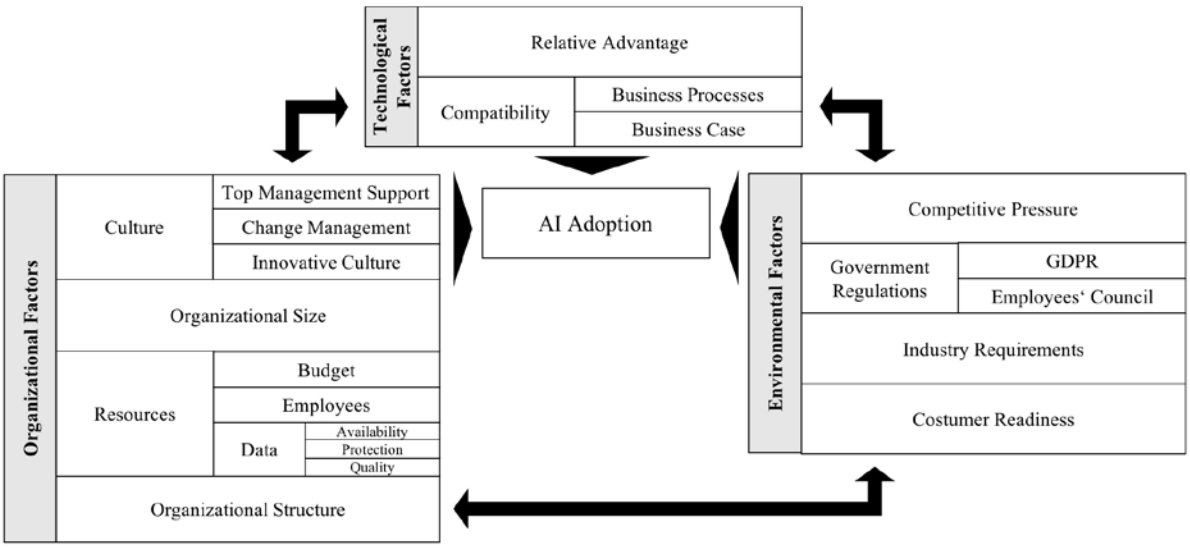

Figure 5. Extended and Deepened Framework for AI Adoption

An example of this misconception between those two groups is the differing assessment of consumers. While user firms tend to view their costumers as sceptical about the acceptance of intelligent applications, providers see consumers as disruptors who explicitly demand innovations.

"I believe we must not forget that our clientele is, to a large extent, rather conservative. And such a chatbot would not be suitable for everyone, not even for half of our target group." – P-02 (UF)

"Very important, I have also become aware of this very often and very clearly, the customers are, as they say, the disruptors. They say exactly how they would like best to work with the brand." – P-10 (PF)

But it is not only the customers that are assessed differently by the respective category of the firm. There is also a divergence of ideas about the prerequisites within the companies. For example, user firms see the size and bureaucracy of their group as an obstacle to the acceptance of AI, while the provider designs their products primarily for large firms in mass markets which can generate sufficient amounts of data.

"Because it has been said that we do not see it within our existing group structures, we cannot give the issue the attention it needs." – P-05 (UF)

"That especially companies that have many service requests benefit from this. [...] I also believe that, for medium-sized companies or something, I do not know. Especially larger companies." – P-08 (PF)

The evaluation of the interviewees' statements also shows that the ideas regarding the availability of budget for AI projects diverge. While large user firms state that they have problems providing the required financial resources, the provider firms overestimate the possibilities of their customers.

"I think obstacles, why we have not done it [AI adoption] yet, are certainly the initial expenditures. At the beginning where you would need a small one-off budget, a bit of know-how as a starting point, which might not be there yet." – P-02 (UF)

"It's also often the case that large corporations in particular have strategic investment pools, where even a CEO says 'yes, I have understood that in order to do something there, we now have to take three, four million in to our hands and we'll take that as play money and start making this initial investment'." – P-14 (PF)

Another point mentioned by the provider firms is the preference of user firms regarding on premise versus cloud-based solutions. As a result, providers are often unable to train and adapt the intelligent algorithms adequately since access to data and sufficient computing power is constrained.

Considering the differences between user and provider firms, we posit the following propositions:

Proposition 9: *The diverging assessment of consumer's AI readiness by provider and user firms leads to a different estimation of demand and thus will have a negative effect on adoption of AI in companies*

Proposition 10: *The fact that the companies that have sufficient data volumes and are addressed by provider firms are also trapped in slow structures of their corporations will have a negative effect on adoption of AI in companies*

Proposition 11: *Misconceptions about budget availability and willingness to pay between user and provider firms will have a negative effect on adoption of AI in companies*

Proposition 12: *Differing preferences of cloud-based and on premise applications between provider and user firms result in a negative effect on adoption of AI in companies*

3.5 Conclusion, Limitations, and Future Research

The explorative study showed that the TOE framework is applicable to the adoption of AI. However, some categories show results that are partially contradictory and require further research (e.g., organizational size). Furthermore, we were able to identify new, AI-specific factors (e.g., data) and subcategories for existing ones (e.g., GDPR and employees' council as part of government regulations). Moreover, evaluating the interviews allowed us to provide initial solution approaches to address the problems that could possibly arise while implementing AI. Altogether, a framework for the adoption of AI is proposed, which provides executives with a broad overview of AI related conditions in organizations. This enables companies to carry out a structured analysis of their status quo and identifying areas of improvements to adopt AI successfully in their processes and services. In addition, it is shown how a gap between supply and demand for AI technology can arise due to diverging assumptions of user and provider firms. In order to enable the top management to address this disagreement, it is necessary to expose them and to create the prerequisites needed for a successful implementation of AI in their company. Besides the practical implications, by conducting the first cross-industry exploratory study focusing on factors which enable and impede AI adoption in general, a basis for further research is introduced. This study can be seen as a starting point to conduct additional studies – for example focusing on or comparing special industries (e.g., healthcare, banking and finance) and associated requirements or looking at specific departments and use cases in depth (e.g., HR, Service).

Future research should consider a constitutive quantitative study, to review the given proposals and further examine existing inconsistencies within the factors. This will help to understand the factors' actual impact, making it possible to develop sound strategies and action plans for an integrated AI adoption. Moreover, a framework other than TOE might then be applied to better reflect the specific requirements of AI (e.g., conceptual framework of organizational innovation adoption by Frambach and Schillewaert (2002)). In addition, companies across the globe and of various cultures, should be included in the research, although a semi-multinational context already exists due to the fact that the interviewed firms are operating in several countries. Additionally, we have mainly considered large companies so far, as they currently already have dedicated positions for AI projects and could therefore be easily identified and contacted. However, future research should survey medium-sized and smaller companies, especially as contradictory results on the impact of company size were obtained in the study. Nevertheless, this study ultimately was able to conceptualize an 'organizational chassis' for the introduction of AI adoption that enables organizations to move forward in the field of AI.

4 Paper B: A Qualitative Analysis of Organizations' Usage of Data Science Competitions

Title

Crowdsourcing Data Science: A Qualitative Analysis of Organizations' Usage of Kaggle Competitions

Authors

Tauchert, Christoph, Technische Universität Darmstadt, Germany

Buxmann, Peter, Technische Universität Darmstadt, Germany

Lambinus, Jannis, Technische Universität Darmstadt, Germany

Publication Outlet

Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS-53), Wailea, Hawaii, USA

Abstract

In light of the ongoing digitization, companies accumulate data, which they want to transform into value. However, data scientists are rare and organizations are struggling to acquire talents. At the same time, individuals who are interested in machine learning are participating in competitions on data science internet platforms. To investigate if companies can tackle their data science challenges by hosting data science competitions on internet platforms, we conducted ten interviews with data scientists. While there are various perceived benefits, such as discussing with participants and learning new, state of the art approaches, these competitions can only cover a fraction of tasks that typically occur during data science projects. We identified 12 factors within three categories that influence an organization's perceived success when hosting a data science competition.

Keywords

crowdsourcing, data science, organization, success

4.1 Introduction

"Data is just like crude. It is valuable, but if unrefined it cannot really be used. […] So must data be broken down, analyzed for it to have value" (Palmer 2006). When companies want to refine their valuable data treasures they face various questions such as: How to deal with large amounts of data? How to extract valuable insights from the data? How can the business benefit most from the utilization of data? To create value from data, companies employ data scientists who analyze the data that the company holds.

According to the 2019 Gartner CIO report, companies are struggling with an acute shortage of talents when it comes to their efforts in implementing artificial intelligence (Costello 2019). Since data science is heavily related to machine learning and therefore artificial intelligence, this shortage also affects the companies' efforts to turn their data into value.

One theoretical possibility to deal with the scarce resource of data scientists could be to leverage the concept of crowdsourcing. The method to draw on the so-called wisdom of the crowd for problem-solving has been established in various domains for several years. Since data science is a fairly new domain, the use of crowdsourcing has not been adopted largely, yet. One platform that enables companies to seek help from a wide range of data scientists is Kaggle.com. The website's focus is hosting machine learning competitions, organized by the respective companies, for which participants try to build prediction models.

While there generally has been a lot of research done for crowdsourcing, there is, after an extensive investigation, almost no research available addressing the combination of both, crowdsourcing and data science. The overall objective of this study is to provide an overview of crowdsourcing in data science, with a special focus on factors that influence the organization's perceived success of a data science competition. To facilitate the achievement of this objective the study uses expert interviews that are conducted with data scientists from different industries. The interview data is enriched with data that is crawled directly from the data science platform Kaggle.

The research questions this study attempts to answer are as follows:

(1) For what purpose do organizations host data science competitions?

(2) Which factors influence the organizations' perceived success when hosting a data science competition?

The remainder of this manuscript is structured as follows: To begin with, we provide a brief overview of the theoretical background and related research to mark off the research area before the qualitative study design is presented. After introducing our study sample comprising ten interviewees, we derive the results. Finally, we conclude the manuscript by pointing out the limitations of our study and providing specific avenues for future research.

4.2 Theoretical Background

4.2.1 Data Science and Kaggle Competitions

In recent years, the term data science has become a buzzword that is surrounded by a lot of hype. An article of the Harvard Business Review even designated data scientist as "the Sexiest Job of the 21st Century" (Davenport and Patil 2012). On the other hand, there are voices, who have criticized the closeness of the definitions of the terms data (or business) analytics and data science, but due to new types of data, new methods and new questions a change in the wording is accepted (Bichler et al. 2017; Dhar 2013). Van Der Aalst defines data science as follows: ''Data science is an interdisciplinary field aiming to turn data into real value. […]. The value may be provided in the form of predictions, automated decisions, models learned from data, or any type of data visualization delivering insights. Data science includes data extraction, data preparation, data exploration, data transformation, storage and retrieval, computing infrastructures, various types of mining and learning, presentation of explanations and predictions, and the exploitation of results taking into account ethical, social, legal, and business aspects''(van der Aalst 2016).

A fundamental concept of data science is to systematically extract useful knowledge from data to solve business problems (Provost and Fawcett 2013). A widely accepted codification of this process is the CRISP-DM (CRoss Industry Standard Process for Data Mining) framework. The entire process is described by six phases on a highly aggregated level (Chapman et al. 2000):

(1) Business Understanding: The purpose of the initial phase is to understand the customer's needs, determine major factors that have to be considered and formulate business objectives.

(2) Data Understanding: The second step consists of data collection, description, exploration, and verification.

(3) Data Preparation: This phase continues with the handling of data by "cleaning" it to be suitable for later analysis.

(4) Modeling: The fourth phase of CRISP-DM starts with the actual selection of the modeling technique. A subset of the data has to be selected for training, testing, and evaluation of the model. Afterward, one or more models are built with varying parameters whose output can be

evaluated. The evaluation is based upon the domain knowledge, the data mining goals chosen in phase one and the test design.

(5) Evaluation: This phase deals with the evaluation of the model with regard to the set business objectives. At this point, it has to be decided whether the model satisfies all requirements.

(6) Deployment: The final phase of the framework addresses the issue of actually deploying the model as well as how to maintain and monitor the outcomes of the project in the long run if used in daily business.

Our study focuses on Kaggle, which is the world's largest online platform for data science with more than 1,000,000 members. While the platform is a large repository for public datasets and a place to exchange for data scientists through discussion forums and public Jupyter notebooks, its main feature is hosting machine learning competitions for various organizations (Metz 2013).

The general concept of a Kaggle competition requires participants to develop a prediction model for a precisely defined problem from given data. The submitted models are evaluated in real-time and the respective prediction score is shown in a leaderboard, which creates a competitive environment. However, the final ranking is calculated based on a separate non-public subset of test data. Afterward, the participants that created the highest-ranked submissions receive the prize money, often in return for the intellectual property of the solution (Kaufmann et al. 2011).

However, when comparing the tasks of data scientists and the concept of Kaggle competitions, it seems that these competitions do not allow to crowdsource all activities related to data science but only a subset. While data science is also about understanding the business, identifying fields of application as well as required and available data, the scope of the competitions only covers tasks closely related to machine learning, like data cleaning and model building.

4.2.2 Crowdsourcing

The idea behind crowdsourcing is that an organization proposes the voluntary processing of a task that is presented in an open call to an undefined group of individuals or teams (Estellés-Arolas and González-Ladrón-de-Guevara 2012). A strength of crowdsourcing lies in the open call to the broader public which can serve as a means to obtain new ideas and approaches from people outside the usual domain and boundaries (Afuah and Tucci 2012). Crowdsourcing can be collaborative or competitive. The former encourages participants to collectively work towards a common solution while the last one aims at the collection of various independent solutions out of which the crowdsourcer can select the winning solutions (Afuah and Tucci 2012). Competitive crowdsourcing initiatives often result in a financial or non-financial compensation of winning participants (Zhao and Zhu 2014). The Kaggle competitions described above fall into the category of competitive (or tournament-based) crowdsourcing.

The broad adoption of crowdsourcing led to a large number of scientific papers examining this topic with various different foci. However, since a crowdsourcing task's success and thereby likewise the overall success of the hosting platform itself, is significantly dependent on the number of individuals participating at a given task, research has focused on the user's perspective of crowdsourcing.

Studies addressing the users' motivation to participate in crowdsourcing usually consider two distinct kinds of motivation, i.e., extrinsic and intrinsic motivation, drawing on the self-determination theory (Deci and Ryan 1985). Extrinsic motivation refers to performing an action to attain an external result. In other words, the incentive is coming from an outside source. Intrinsic motivation, in contrast, is independent of some outcome but arises from the pure fun and joy of doing something (Deci and Ryan 1985). The incentive is to be satisfied and the task itself is central instead of a promised reward (Ryan and Deci 2000). Factors of motivation that were identified include: task autonomy and skill variety as factors of fun and delight (Kaufmann et al. 2011; Pilz and Gewald 2013; Zheng et al. 2011), financial compensation (Kaufmann et al. 2011; Leimeister et al. 2009; Pilz and Gewald 2013; Zheng et al. 2011), social motivation (i.e., reputation) (Hertel et al. 2003; Kaufmann et al. 2011; Pilz and Gewald 2013), tacitness (Zheng et al. 2011), learning (Hertel et al. 2003; Leimeister et al. 2009), self-marketing (Leimeister et al. 2009), meaningfulness / impact of the task (Chandler and Kapelner 2013), complexity (Eickhoff 2014; Shao et al. 2012; Zheng et al. 2011), event duration (Shao et al. 2012), number of events (Shao et al. 2012).

From an organization's perspective, crowdsourcing is designed to get others to solve problems by using knowledge that the organization may not normally have access to (Jeppesen and Lakhani 2010). And therefore, the main reason for organizations to initiate a crowdsourcing campaign is to get the result of a given task or the resolution of a problem (Estellés-Arolas and González-Ladrón-de-Guevara 2012). Often crowdsourcing is associated with innovation processes such as new product development or product improvements (Poetz and Schreier 2012). In this case, companies get creative ideas, that might be commercially exploitable (Kleemann et al. 2008). This approach is supported by studies that show that many of those user innovations are characterized by high commercial attractiveness (von Hippel 2005).

Besides concrete innovations, companies also try to create any type of added value by crowdsourcing through value creation or increased profits (Yang et al. 2008). Another goal that organizations might pursue through crowdsourcing is to obtain knowledge and especially talent from the crowd by using crowdsourcing campaigns as an employee recruitment tool (Howe 2006).

We found one study that used Kaggle as a context (Garcia Martinez 2015). It assessed how participants' engagement is related to their solutions' creativity. The results show that higher cognitive and emotional engagement is associated with more creative output. Further emphasis is put on the willingness to share obtained knowledge. The data shows that the need for versatile problem-solving skills makes a competition intrinsically inspiring, which in turn strengthens the desire to share a promising solution with others.

To summarize, so far a lot of research on crowdsourcing has focused on the motivation of users to participate in and companies to host crowdsourcing events. The present study aims to give insights into the organizations' perspective on the success of data science competitions. Therefore, it provides a basis to fill the research gap that currently exists in this area.

4.3 Method

The goal of our study was to expand the current stage of IS research concerning the crowdsourcing of data science projects. Since the amount of companies that are conducting data science challenges on platforms such as Kaggle is low and the field has not been extensively explored, an explorative approach using interviews with experts seems appropriate to investigate the problems occurring in this particular context (Flick 2004). According to Weber (Weber 1990), content analysis is an appropriate approach to assess open-ended questions and therefore, it is suitable for the evaluation of the collected qualitative data. The interviews were transcribed, coded and analyzed taking into account relevant publicly available data through triangulation (Hsieh and Shannon 2005). Therefore, we collected data from Kaggle using a self-written crawler and conducted explorative data analysis. The data was crawled mid-December 2018. We decided to use Kaggle as a context since it is by far the largest (most registered and active users) and most open (commercial and non-commercial) platform for data science competitions. The two alternative platforms, Codalab and RAMP, are intended for research problems only.

4.3.1 Research Design

Our main information source was in-depth expert interviews, which were conducted in a semi-structured way. Following the guiding principles of Sarker et al. (2013), we prepared an interview protocol and acquired key informants in different companies using professional social networks (i.e., LinkedIn, XING). During the interviews, we kept our questions open in order to enable participants to speak freely.

The interview guide comprised five different sections: The first part comprised general questions about the interview partner and the company he/she works for and introduced the context of the interview. The second section tackled the topic of how data science is used in the company in general. In the third section, we focused on data science competitions and asked the interview partner about their experiences and opinions about data science challenges on internet platforms. In the fourth section, we wanted to know how data science platforms, in general, are perceived by the experts. In the last section, the informants had the chance to comment openly on the topic and add remarks.

Due to the semi-structured approach, questions were gradually adjusted in order to account for the interview partners' individual situation.

4.3.2 Sample and Data Collection

Table 4 provides an overview of the participants. The interviews were conducted over a three-month period and took place between November 2018 and January 2019. In total ten interviews with highly involved participants were conducted of whom all were data scientists. After the tenth interview, data collection was discontinued since no new, previously unmentioned aspects were mentioned (Flick 2004).

The average duration of the interviews was approx. 30 minutes and the interviews were mostly held via telephone due to geographical distance.

ID	Industry	Total Employees	Data Scientists	Revenue [bn. €]
A	Telco	10,000 - 100,000	10 – 50	> 20
B	Research	< 10,000	-	-
C	Government	< 10,000	< 10	-
D	Financial Services	> 100,000	10 – 50	> 20
E	Chemical	10,000 - 100,000	> 50	5 - 20
F	Research	-	-	-
G	Software	10,000 - 100,000	> 50	> 20
H	Price Comparison	< 10,000	< 10	< 5
J	Automotive	> 100,000	> 50	> 20
K	Financial Services	-	< 10	-

Table 4. Sample Description

We used a conventional approach to content analysis, which aims to describe a phenomenon to allow new insights to emerge (Hsieh and Shannon 2005). This is also described as inductive category development (Mayring 2004). Subsequently, the transcripts were assessed by using the MAXQDA software and by conducting two coding cycles as recommended by Saldaña (2015). The first coding cycle comprised a mixture of attribute coding and descriptive coding. The former was performed to obtain essential insights about the data and its descriptive information. The latter was used to extract additional aspects, key thoughts, and concepts from the interview data. In a second cycle, the formerly created codes were combined into a smaller number of sets using pattern coding (Saldaña 2015). By discussing and assessing the coding process with a group of three IS researchers and students, an investigator triangulation helped to ensure rigor and trustworthiness. Furthermore, the crawled data from the Kaggle platform was used for data triangulation (Flick 2004).

4.4 Results and Discussion

4.4.1 Information about Competitions

Competitions constitute the most important aspect of Kaggle since it is the service it started with and it is still mainly what they are known for.

Until the date of data collection, 309 competitions have been hosted at the platform, including the still-active ones. 50.5 % of competitions were hosted by companies or organizations, which provided an explicitly defined problem to be solved and offered a reward (mostly price money). 24.6 % (76) competitions were categorized as *research*. The entities behind those competitions are usually non-commercial institutions with some scientific background. They are thus often not able to provide as much prize money as commercial companies. To facilitate research competitions Kaggle offers to sponsor them by providing \$25,000 as prize money. 16 competitions (5.2 %) were in the category "recruitment". In general, these competitions do not differ from the aforementioned competitions except that they offer job interviews for the highest-ranking participants. The other competitions belonged to the categories *playground*, *getting started*, *masters* and *analytics*.

Companies planning to host a competition have to compete with other active competitions for the attention of users. One factor that can be directly influenced by the firms and that might increase Kagglers' motivation to participate is the prize money rewarded to the highest-ranking participants. Table 5 shows the statistics for rewards and participants for competitions that offered any prize money (> \$0).

Statistic	**Reward**		**Teams**	
	Featured	Research	Featured	Research
Count	154	61	154	61
Mean	49,219	7,051	1,128	320
Max.	1,200,000	25,000	7,198	1,386

Table 5. Rewards' and Participants' Structure of Competitions

The numbers represent 215 competitions in total, thereof 154 featured and 61 research competitions. It can be seen that the mean prize money for category featured (\$49,219) is seven times the amount for research competitions (\$7,051). Regarding the number of participating teams per competition, we see that featured competitions (1,128) have about 3.5 times as many teams as those with a research label (320). One reason for this might be the in average significantly lower prize money. Another reason might be that the Kaggle community is more interested in industry competitions than in research competitions.

The USA has hosted the majority of competitions, accounting for 65 % of all competitions with 138 hosted competitions. About 34 competitions are coming from companies and research institutions in Europe, mainly from the United Kingdom (9), France (8), Spain (7) and Germany (6). Asian countries with participating companies are mainly Russia (5), Japan (5), Israel (3), Taiwan (2) and China (2).

4.4.2 Results

While coding the transcribed interviews, we noticed that codes could be categorized in: (1) platform-related, (2) organization-related and (3) outcome-related factors.

Platform-related factors

Community. The capabilities of the data scientists on Kaggle are considered to be very high. Hence, several experts (A, C, J, and K) see a good chance to obtain high-quality models from a competition. With an average of 1,128 teams that are participating in such competitions the potential for great ideas and solutions is relatively high. Experts C and J say that Kaggle competitions attract some of the best data scientists in the world, such as for example Tianqi Chen, the lead developer of the popular XGBoost framework, participated at eight competitions. In addition, experts F and K perceive their respective competitions as successful, even though their competitions are not finished at the time of the interviews. They are both largely satisfied with the number of participating teams, as it means potentially a lot of new ideas (for expert F) as well as many people being aware of the company, who wants to increase brand popularity (for expert K). With having two to three times more people than expected and reaching the targeted number of 1,000 teams within the first week, respectively, it is apparently relatively easy to attract a lot of people and motivate them to participate. These statements correspond with the data retrieved from Kaggle showing an average of about 1,100 participating teams per competition. A possible reason for such a high number might be that the number of new competitions is not steadily growing, as one might suspect, but is instead staying at a relatively constant level of about three new competitions per month. Expert E mentions that an ambitious participation at a competition is accompanied by an expenditure of time close to full time. Therefore, it can be assumed that Kagglers, in general, do not participate in multiple competitions simultaneously. More simultaneously active competitions would thus reduce the average number of participants per competition, which would be counterproductive as companies try to attract as many Kagglers as possible. A study of Shao et al. from 2012 supports this presumption. The study's findings suggest that a higher competition intensity in a crowdsourcing context is associated with a significant decrease in participating users (Mayring 2004).

Infrastructure. By providing data storage capacities for data sets and computing power for machine learning models, Kaggle is removing barriers that would otherwise hamper companies to organize such data science competitions. Companies struggle enough with the collection and preparing of data and therefore are happy that they do not have to worry about technical infrastructure.

Regulations. While Kaggle is trying to have a low technical barrier, they do have other barriers in place. The minimal amount of prize money for *featured* competitions is $25,000. Depending on the company size that might be a lot of money to spend on an unknown outcome. Especially small and medium-sized companies, who could really benefit from this approach, could be scared off for this reason. Another restriction Kaggle imposes on the hosting organization is, that only supervised machine learning problems are allowed. Companies whose field of activity is in an area where unsupervised or reinforcement learning approaches are necessary cannot host a competition. Expert F and her team started the first competition with the intention to have it as the first of a whole series of competitions. Since her team is especially interested in unsupervised learning problems such as anomaly detection, they are reconsidering whether they complete the series of competitions.

Organization-related factors

Marketing. A further reason for experts H and K to host competitions is to do public relations or brand building. The experts say that the proper utilization of a company's data is getting more and more crucial to stay competitive in the business. The market for data scientists, however, is very small. It is therefore important that data scientists, as potential employees, know the company and recognize the brand. By hosting a machine learning competition, the firms try to increase their attractiveness towards the data science community. Another means of marketing are hackathons. Those originally from software development coming short-term events have been named by several experts (A, B, D, and H) when they have been asked if they have hosted a machine learning competition so far. The association between a machine learning competition on Kaggle and a hackathon can be seen as reasonable, as hackathons usually do not create fully-fledged solutions but rather partially usable prototypes and furthermore are intended to increase the brand awareness among possible employees or customers (Frey and Luks 2016). This coincides with what the experts think about the results of Kaggle competitions, as mentioned above, and also with the aim to engage in brand building. One obvious difference, however, is that hackathons are local in general, while a Kaggle competition reaches out to a worldwide distributed audience. It can, therefore, be assumed that Kaggle is a new means of marketing to reach out to the data science community and complements the established practice of hackathons.

Recruiting. One of the incentive types on Kaggle is the prospect of a job interview at the hosting company. As there is a high demand for data scientists, it seems to make sense to draw on a data science community as large as Kaggle to get in touch with potential employees. However, the data analysis shows that the competition category *recruitment* has only been chosen 16 times, with the last appearance in the first quarter of 2017. These findings suggest that companies do not like this option, maybe because it did not prove to be successful. The three interviewed hosting companies (B, F, and K) are in line with this development and do not focus on recruitment through Kaggle. Expert F, who is working for a research institution, says that for an academically career other skills are higher valued than those skills that can be shown at a competition. In addition, expert K, who is working for a commercial company, states that recruitment would be a nice side effect but not of special interest. The experts E and G have used the Kaggle job board successfully in the past, which is not directly linked to the competitions but presents regular job advertisements. Expert K additionally mentions that an advantage of a featured or research competition is the participation of a worldwide-distributed audience. Although a recruitment competition is in general free for everyone to join, she might be right because a certain proportion of potential users might not be motivated to participate, assuming a job offer is unappealing for participants not looking for a job. A lower participation rate, however, would have a negative impact on the important objective of obtaining new ideas and innovative approaches from submitted solutions. This statement is in line with expert J, saying that about 70 % of a data scientist's actual work is not required on Kaggle. In the remaining 30 %, however, participants can excel and obtain excellent knowledge, according to him. Expert K emphasizes that the participants' aim on Kaggle is always to get a high final score, i.e. to maximize the accuracy of the model, whereas in a real-world problem other aspects such as the speed of a prediction or interpretability might play a major role.

Data. Seven out of the ten experts that have been interviewed are working for companies that have not been hosting a competition on Kaggle yet but are considering it (experts A, C, D, E, G, H, and J). When asked for reasons that might justify this, often their first answer was the apparent need to publish sensitive data. For most companies, a problem that theoretically would be suitable to be solved through crowdsourcing, would contain some type of sensitive data, be it internal data about the company and its projects or customer data, which would potentially allow identification of those customers. Although there are possibilities to anonymize data (e.g. k-Anonymity (Bayardo and Agrawal 2005) and L-diversity (Machanavajjhala et al. 2006)) the companies apparently shy away from putting the sometimes considerable amount of effort into it. As those methods also cannot fully guarantee that any identification can be ruled out (Li et al. 2007), they might not want to take the risk of having a public data scandal. Experts E and G mention that their companies' conservative attitude towards sensitive data-related projects in public is typical for German companies. Research has shown that there are differences in the innovation and risk culture between for instance the United States of America and Europe, with European cultures being more reserved (Ezell and Marxgut 2015). The experts' opinion corresponds with the findings of the data analysis regarding hosted competitions, as about 65 % of all competitions are hosted by US-based companies or institutions, even when competitions hosted by Kaggle and Google itself are excluded.

Top Management Support. Only expert E states that the decision-makers in his company presumably do not know about the possibilities of crowdsourcing for data science projects. However, according to him, this would be the most significant factor why no competition has been hosted so far. Therefore, it seems that the awareness of data science platforms, in particular Kaggle, is fairly high among decision-makers working in the realm of data science.

Use Case. Additionally, expert E as well as expert A say that they did not have any problem that they wanted to get solved by the crowd. At this point, it remains unclear whether they have all the necessary resources to solve the problems internally to a satisfying extent or whether they do not have problems suitable for a Kaggle competition, which are only supervised learning problems so far. However, it seems to be unlikely for companies of their size (both 10,000 to 100,000 employees and revenue of at least $5 bn. per year) to not have any business problem linked to supervised learning.

Lack of Resources. However, for expert B the further usage of Kaggle is less dependent on the features provided by the platform but more on how the competitions are organized within his institution. He, as well as expert K, states that they did all the work of hosting the competitions in parallel with their regular full-time job. The expert, therefore, would prefer to have a dedicated team working on the task of organizing, conducting and evaluating the whole competition to increase efficiency, which so far is not the case.

Outcome-related factors

Innovation. Independent whether their company has been active on Kaggle or not, all experts do name the innovative power behind the competitions as a decisive reason. The two words "new ideas" spring up regularly during the interviews, although no question specifically asks for it. The capabilities of the data scientists on Kaggle are considered to be very high. Hence, several experts (A, C, J, and K) see a good chance to obtain high-quality models from a competition. As stated before, Kaggle competitions attract some of the best data scientists in the world.

Incompleteness. Interestingly, none of the three experts working for a hosting company expects to receive a fully completed machine learning model. Although expert K hopes for a high-quality model, she does not take it for granted and expresses herself cautious about the upcoming results. The two other experts (B and F) do not even expect a solution, which is able to solve the respective problem. Instead, their plan is to closely examine submissions for different approaches on how to tackle their problems. They hope to see approaches that their team did not think of but that show promising results. This way of thinking is presumably found rarely on other established crowdsourcing platforms, e.g. Amazon Mechanical Turk or 99designs, where actually usable and finished results are expected in general. However, the differences in the complexity between the tasks on those platforms compared to tasks on Kaggle are considerably high, making a direct comparison difficult. The concept of using the community for solution finding is closely related to "open innovation", where companies integrate external sources into the usually internal innovation process. The external sources get reached via an open call to a large, unknown crowd (Chesbrough et al. 2006). This is very similar to the definition of crowdsourcing. Open innovation is not intended to replace but to complement the traditional innovation process (Chesbrough et al. 2006), which is in line with the statement of expert B, saying that crowdsourcing in data science is not used to replace the internal process but used as an additional channel. Kaggle, therefore, seems not so much to be about actually solving a problem directly but to support the organizing company at ultimately achieving a complete solution.

Learning. Expert F sees high value in monitoring the progress of participants through closely following the discussions on the competition forum and in answering those questions. As most user presumably do not have the same domain/business background as the organizing team, they approach the problem unbiased, which includes interesting information for the team of expert F. The data analysis verifies that there are indeed a lot of discussions during a competition with an average of 101 threads per competition. Considering that the average competition lasts for 78 days, this means more than one new thread per competition and day. The expert's statement shows that the crowdsourcing process on Kaggle is not just done by providing a relevant problem with subsequent waiting for a fitting solution, but that it is more a constant, interactive and collaborative process with learnings on both sides. The assessment of the overall success of a competition is therefore not solely dependent on the best final solutions but also on the process to reach them.

Figure 6 depicts all identified factors and their influence on the organization's perceived success of data science competitions.

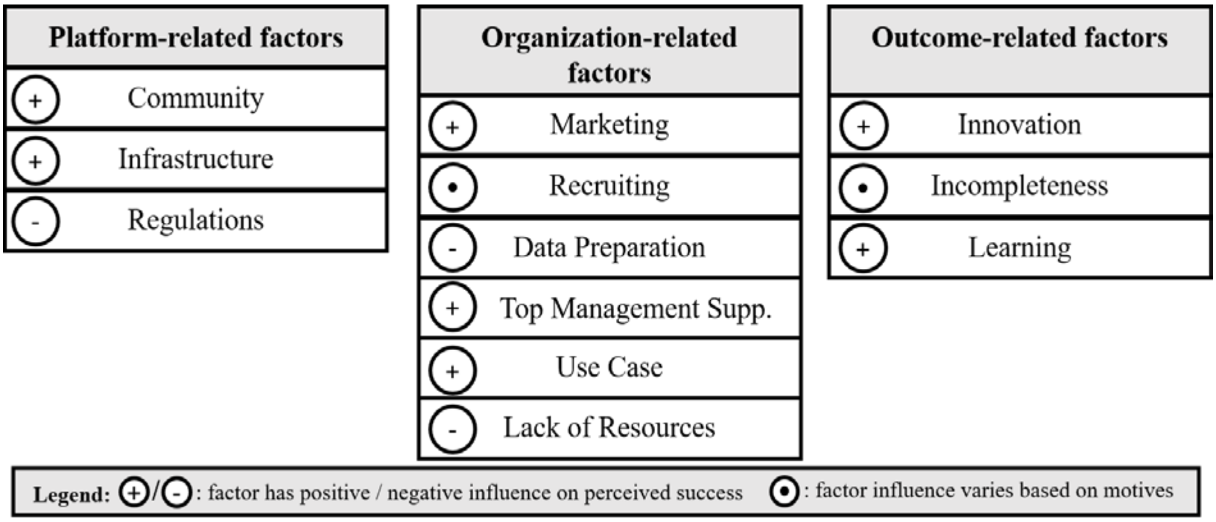

Figure 6. Factors Influencing the Success of Data Science Competitions

4.5 Conclusion

The combination of crowdsourcing and data science is a relatively new concept, which has not been exhaustively researched. Therefore, this study creates a basis for further studies in this context. We enriched the qualitative interview data with data that we crawled directly from the Kaggle platform. This approach allows for a broad overview of different interesting aspects and data triangulation.

The data shows that so far 32 companies and research institutions have hosted at least two competitions, some of them up to four. It, therefore, seems that for some companies crowdsourcing in data science might indeed work and deliver good solutions.

The interviews show that companies highly value the innovative power of the community of data scientists on Kaggle but see problems in dealing with sensitive data in a public context. Brand building and partially recruiting are seen as positive aspects.

Crowdsourcing proved to be a valuable concept for companies to leverage the wisdom of a heterogeneous crowd. Data science is currently rapidly expanding and still in a relatively early stage. The combination of both fields promises a lot of potential. As more companies and interested people get in touch with data science, platforms like Kaggle might emerge creating a competitive market. A lot of research needs to be done to obtain further insights into this new market comprising the combination of crowdsourcing and data science.

The present study examines the relatively new combination of crowdsourcing with data science. So far there has been almost no research conducted in this specific context. This explorative study aims to serve as a basis for further studies in the context of crowdsourcing in data science.

The main reason for companies for hosting a machine learning competition is the innovative power inherent in the wisdom of the crowd. It is important to obtain insights, whether the solutions, especially the winning ones, actually deliver the desired innovation. Therefore, further research should, among other things, focus on companies that have hosted competitions in the past, which also means that another cultural context will have to be considered. The other part of companies, i.e. those who do not host competitions, see the biggest problem in the publishing of sensitive data. It is important to know how rational this justification actually is, and how well common anonymization techniques can be utilized to make datasets suitable for those competitions. Thereby, companies could better assess the risk related to hosting. As marketing reasons are also named by the experts, research should get insights about the actual perception of companies in the community. Ultimately, it needs to be examined whether machine learning competitions are indeed an appropriate marketing tool to increase brand awareness in the data science community. Furthermore, it is essential for companies to know how to design a competition, e.g. in terms of prize money, duration as well as topic and problem description, respectively. Therefore, a closer comparison between more and less successful competitions is needed.

The results of this study indicate that crowdsourcing and data science can be combined in a successful manner. However, companies, which plan to host a machine learning competition, should bear in mind that the circumstances are appropriate. Firstly, in most cases, Kaggle is presumably not a way to get a given problem solved by others for cheap money in a short time. Rather, the crowd should be seen as a means to enrich the internal data science process. Permanent communication and collaboration between participants and the host are most likely to be the best way to achieve promising results. To ensure such a process, companies should provide a dedicated team of internal employees to organize and supervise the competition instead of doing it next to daily work. If companies have a well-designed backend system for their data, which allows for easy preparation of datasets, Kaggle is more likely to serve as a good platform to use the wisdom of the crowd for problem-solving. Otherwise, composing a well-suited dataset can be a difficult and time-consuming task.

As every study, also the present study and its results are to be seen and interpreted in consideration of certain limitations. Since this study is based on a relatively small sample of only ten interviews, we cannot draw confident conclusions. Furthermore, this study aims to provide broad oversight of the subject matter. Therefore, the different aspects are examined at a very high level and are only scratched on the surface. The experts' answers in the interviews are naturally at least partially subjective and should not be seen as a matter of fact. Lastly, with only three experts working for hosting companies, the generalizability of their answers needs to be evaluated carefully.

5 Paper C: Promoting Trust in AI-based Expert Systems

Title

Promoting Trust in AI-based Expert Systems

Authors

Mesbah, Neda, Technische Universität Darmstadt, Germany

Tauchert, Christoph, Technische Universität Darmstadt, Germany

Olt, Christian M., Technische Universität Darmstadt, Germany

Buxmann, Peter, Technische Universität Darmstadt, Germany

Publication Outlet

Proceedings of the 25th Americas Conference on Information Systems (AMCIS 2019), Cancun, Mexico

Abstract

Recent advantages in artificial intelligence (AI) research allow building sophisticated models to advise users in various scenarios (e.g., in financial planning, medical diagnosis). For companies, this development is relevant since it allows scaling of services that were not scalable before. Nonetheless, in the end, the user decides whether he/she uses a service or not. Therefore, we conducted a survey with 226 participants to measure the relative advantage of AI-based advisory over human experts in the context of financial planning. The results show that the most important advantage users perceive is convenience, since they get easy and instant satisfaction of their informational needs. Furthermore, the effectivity of eleven measures to increase trust in AI-based advisory systems was evaluated. Findings show that the ability to test the service noncommittal is superior while the implementation of human traits is negligible.

Keywords

artificial intelligence, expert systems, promoting, acceptance, robo-advisory

5.1 Introduction

What do you do when you are on the lookout for information on how to invest your money? Do you make an appointment with your bank advisor? Do you gather information on investment opportunities on one of the most recent fin-tech websites? Or are you already using a financial robo-advisor – a service that provides financial advice or management with minimal human involvement by interacting only with an information system? If you choose the latter, then you belong to a minority of roughly 45 million people who have invested money using a financial robo-advisor (Statista 2019). However, intelligent algorithms that provide advice are not an innovation solely relevant for the financial sector. To give an example, in medicine, real estate or insurance, artificial intelligence (AI) is changing the decision-making process right now.

Robo-advisors are automated advisory services. Customers are guided through a self-assessment process and get a target-oriented recommendation (Jung, Dorner, Weinhardt, et al. 2018; Sironi 2016). By using AI-based algorithms, robo-advisory systems can process and utilize more information than any human while at the same time they are cheaper and superior in terms of scalability compared to human experts (Tertilt and Scholz 2017). Despite these existing advantages, the use of AI-based advisory systems in enterprise-client interaction has not yet become established (e.g., Jung and Weinhardt 2018). Therefore, we explore the reasons why the use of AI-based advisory services has not yet been adopted and how we can promote the future adoption of using these services. There are two key factors for the adoption of an AI-based expert system. On the one hand the user needs to perceive a true advantage, when he/she use an AI-based advisory service (Choudhury and Karahanna 2008). On the other hand the user needs to trust the AI-based advisor's suggestion (Lin 2011; Pavlou 2018). However, to the best of our knowledge these two factors have not yet been discussed in the Information System (IS) literature. Therefore, our first research question arises:

RQ1: *Do users of expert systems currently perceive the superior capabilities of AI-based advisors compared to human advisors?*

Regardless of whether and to what extent the advantages of AI-based advisory systems are perceived, companies try to influence their customers' perceptions and beliefs. One possibility is to increase trust in AI-based systems. As mentioned before the user's trust is a key factor in innovation adoption (Lin 2011; Pavlou 2018) as well as in following advice (Sniezek and Van Swol 2001; Van Swol 2011). While there is various research on opportunities to increase trust in expert or recommender systems, research usually focuses on one single aspect at a time, e.g., exclusively transparency (Nilashi et al. 2016) or exclusively anthropomorphism (de Visser et al. 2016). We found no studies comparing different mechanisms and assessing their effectiveness in creating customer trust. However, companies cannot put all existing alternatives into practice due to constraints regarding time, money and technological possibilities. They need to evaluate and prioritize between different available options and implement the most effective ones. This leads to our second research question:

RQ2: *Which mechanism that establishes trust in AI-based advisory systems comparatively generates the highest level of trust in AI-based advisory systems?*

To answer these questions, we will first provide an overview of the theoretical background related to relative advantage, trust in technical systems and advice in general. Based on the findings, we will derive hypotheses, which are then tested by means of an online survey among potential users of expert systems. We will present the study design and the study sample of 226 participants before analyzing the collected data using group comparison. Finally, by discussing the findings, we illustrate the implications and will then conclude the manuscript by pointing out the limitations and identifying areas of future research.

5.2 Theoretical Background

Recent advantages in AI research allow creating sophisticated models that are able to leverage vast amounts of data as well as understand and interpret spoken and written human language (e.g., Alexa, DeepMind, IBM Watson). AI is a "science and engineering of making intelligent machines, especially intelligent computer programs", which tries but is not limited to simulate human intelligence and which includes underlying technologies like machine learning, deep learning and natural language processing (Elliot and Andrews 2017; McCarthy 2007, p. 2). This technology is an opportunity for innovations in the advisory industry (Jung, Dorner, Weinhardt, et al. 2018): It enables human experts to be replaced by AI-based systems.

When it comes to the adoption of innovations according to the theory of innovation diffusion, the perceived relative advantage (RA) is the main predictor of innovation adoption (Choudhury and Karahanna 2008). That means customers compare the benefits of using an innovation such as lower costs, less time effort, higher convenience to the status quo and if she perceives a net benefit, she is likely to adopt the innovation (Rogers 2003). Based on Rogers' theory, Choudhury and Karahanna (2008) identified three dimensions of relative advantage of electronic channels: **trust**, **convenience**, and **efficacy** of information acquisition. Trust in this context is framed as institutional trust, i.e., trust in the concept of robo-advisors (McKnight et al. 2002). Hereby, two aspects can be distinguished: (1) informational trust, i.e., the user's belief about reliability and accuracy, and (2) structural assurance, i.e., the user's belief in the technological foundations of robo-advisors (Choudhury and Karahanna 2008; McKnight et al. 2002; Wang and Benbasat 2005). E-commerce users are generally convenience-oriented and try to save time and money as well as an easy way of completing online transactions (Devaraj et al. 2002; Li et al. 1999). The last aspect is the efficacy of information acquisition, meaning that the efficacy of the source of advice (human vs. algorithm) as a medium must fit the equivocality of the information being communicated (Daft et al. 1987). These dimensions are then combined to form factors of relative advantage during different phases of the advice process. The first factor, **RA-Learning** consists of all three dimensions during the phase of learning about the advice context (e.g., financial investment). The second factor, the factor **RA-Informational Trust** includes statements that refer to the confidence in the information that the robo-advisor has provided. The third factor **RA-Informational Convenience** covers statements that are related to the convenience of obtaining information during the advisory process. The last factor, **RA-Transaction** incorporates all statements related to the convenience and confidence in a transaction through a robo-advisor.

When considering the advantages and disadvantages of robo-advisory in terms of perceived relative advantages, the following is striking: (1) Advice from robo-advisors is often given without explanation or without the opportunity to understand why it is given by the robo-advisor (De Laat 2018). With human advisors, interaction is easily possible and their advice can be understood. This disadvantage of robo-advisors could influence the trust dimension of RA. Hence, users of an AI-based advisory system would not perceive a RA compared to human experts. (2) Robo-advisors are always available. It is not necessary to make an appointment like with a human advisor. This advantage could have an impact on the convenience dimension of RA, with the result that based on this dimension, users of an AI-based advisory system would perceive a higher RA compared to human experts. (3) Robo-advisors can process much more data than human experts and therefore they have a larger knowledge base. To gain access to this knowledge base through human advisors, several human consultants would need to be involved. This advantage could have an impact on the efficacy dimension of RA, hence based on this dimension, users of an AI-based advisory system would perceive a higher RA compared to human experts. We expect that especially in situations where a concrete explanation of the advice is not crucial to put it into action, the advantages of a robo-advisor over a human advisor outweigh. The following hypothesis emerges:

***H1:** When using AI-based advisory systems, users perceive relative advantages over human experts.*

When it comes to investigating how a person utilizes advices that he or she gets from another person the Judge-Advisor system (JAS) is an often used and well-studied paradigm (Sniezek and Buckley 1995). Originating in behavioral psychology, it describes a structured group in which one group member, the judge who has the sole power to make a decision, seeks out advice from one or more advisors (Van Swol 2011). In many studies a large variety of factors that influence to what extent the judge incorporates the advice during the decision-making process have been investigated (Bonaccio and Dalal 2006). Examples of such factors are competence, distance of advice, the power relation between judge and advisor (Bonaccio and Dalal 2006; Schultze et al. 2015; Sniezek and Buckley 1995; Van Swol and Sniezek 2005; White 2005). However, researchers agree, that one of the most influential factors of advice utilization is the trust that the judge has in the advisor (Jungermann 1999; Sniezek and Van Swol 2001). Therefore, we investigate how to increase trust in the advisor. With this regard, we were able to identify multiple mechanisms to increase this trust.

Trialability. According to (Rogers 2003), users will feel more comfortable with a product or innovation and are able to give meaning to it, if they can experiment with it. Thus, if consumers are able to **test** the robo-advisor, concerns and perceived risks may be reduced and the advisor's competence and integrity can be confirmed. As a result, users are more likely to trust the advisor.

Anthropomorphism. According to the phenomenon of automation bias, people attribute higher levels of initial trust, higher authority and higher performance expectations to machine-like agents (Dzindolet et al. 2003). Other studies (de Visser et al. 2016) have shown that anthropomorphism might have a positive impact on trust (e.g., when it comes to repair trust). Two possibilities to humanize robo-advisors are (1) to give the robo-advisor a **human appearance** (Hegel et al. 2009), e.g., through a figure, or (2) to let the interactions with the user take place in the form of human-like **dialogue** (Gnewuch et al. 2017).

Transparency. When people get further explanations and gain an understanding of the advice being given, they are more likely to follow advice (Gönül et al. 2006; Zanker 2012). However, transparency in advice taking and decision support systems can have different meanings. First, transparency can be defined as **providing explanations** of the reasons why that advice was given (Nilashi et al. 2016). The concept of explainable AI is a major stream in AI research (Ribeiro et al. 2016). Another aggregated form of explanation can be provided by many algorithms in the way of a **confidence** as probability measure which describes how certain the robo-advisor is concerning its advice (Jung, Dorner, Weinhardt, et al. 2018). Second, transparency can entail the explanation of how the system itself works (Tintarev and Masthoff 2007). In contrast to common software development, it is much harder to implement mechanisms like testability or auditability for AI models, since the code is not based on explicit rules. Instead, the system's knowledge evolves over time. Consequently, information about the used **database** (Jung, Dorner, Weinhardt, et al. 2018), the **technical functionality** (Lipton 2016) as well as how often the robo-advisor was **trained** (De Laat 2018) can create transparency and thus strengthen confidence in the robo-advisor.

Subjective Norms. It is well studied that the opinions of our social environment have a great influence on our behavior (Ajzen 1991; Venkatesh et al. 2003). Therefore, the **recommendation to use a robo-advisor by friends and acquaintances** can have a positive influence on the trust in robo-advisors.

Experience. People often use certificates and awards (e.g., Fund Manager of the Year) as proof of expertise. These are usually earned when individuals are performing well over a period of time. When people acquire several certificates or awards, it tells people that they are constantly performing well and that it is recognized by his/her peers. Feng and MacGeorge (2006) have shown that expertise influences the receptiveness to advice. Information about **previous activities and results** of the robo-advisor (Eule 2017) and **how long the robo-advisor has been in use** (Eule 2017) show the previous experiences of a robo-advisor. These experiences can be perceived as expertise and increase trust in the system.

Summarizing this section, we identified eleven trust-increasing mechanisms: *Testing, Visual Appearance, Dialog, Reasoning, Confidence, Data Transparency, Technical Functionality, Training Frequency, Social Environment, History, Usage time*.

As described above, AI research places a great focus on the topic of reasoning. This is also reflected in the media, where the reasoning of algorithms is demanded repeatedly. Furthermore, lack of reasoning was identified to negatively affect the trust dimension of RA. By strengthening the reasoning capabilities of robo-advisors this disadvantage could be counteracted. Thus, it can be assumed that reasoning has the greatest expected effect on confidence in AI-based systems. Therefore, we hypothesize:

H2: *Reasoning is the most important trust-increasing mechanism for AI-based advice systems.*

5.3 Research Method

To answer the research questions and validate our hypotheses, we set up an online survey that weights the influence of the formerly mentioned factors on trust in AI systems. Therefore, we decided to use robo-advisory in the financial sector as our survey context. Financial robo-advisors are automated investment advisory services. Customers are guided through a self-assessment process and are then recommended a target-oriented investment strategy (Jung, Dorner, Glaser, et al. 2018; Phoon and Koh 2018; Sironi 2016; Tertilt and Scholz 2017). We have chosen the robo-advisory context based on six reasons: (1) Robo-advisors differ significantly from the previous financial planning tool because they take human advisors out of the advisory process (Jung, Dorner, Glaser, et al. 2018). This shows that the use of robo-advisors is a great innovation and a major competitor for traditional financial advisory services (Winnefeld and Permantier 2017). (2) Despite the topicality and novelty of the topic, most people are familiar with the context and the use of robo-advisors is well conceivable for participants (Beketov et al. 2018). (3) The implementation of robo-advisory is very interesting for providers because it offers many advantages, such as the reduction of investment costs or easily scalable service management (Tertilt and Scholz 2017). The potential of the advantages is also demonstrated by the number of start-ups established in this area (Goeke 2016). Many large and traditional companies (i.e., banks) have also identified automatic advisory as a key component of their future strategy (Trentin et al. 2012). (4) There are also some advantages for investors using robo-advisors, such as real-time portfolio surveillance or reduction of investment costs (Tertilt and Scholz 2017). (5) Despite the advantages for providers and customers, the use of financial robo-advisors is still very low (Jung and Weinhardt 2018). Therefore, it is worth investigating how to increase trust in financial robo-advisory systems in order to increase their acceptance and use. (6) Since financial investment is a field where failure can have long-term negative effects, major factors influencing the adoption are different types of risks, i.e., performance, financial, social, security (Lee 2009). The consumer's perception of uncertainty and risk can be counteracted by increasing trust in the concept of robo-advisors (Choudhury and Karahanna 2008). Therefore, increasing trust is absolutely critical in this scenario.

Since we wanted to address a representative and diverse target group of European Internet users in terms of age, gender and employment status, we chose to acquire our participants by using a market research company. Each participant received 0.50€ as incentive. First, we informed our participants that they participate anonymously in a scientific survey and that there were neither correct nor incorrect answers to counteract common methodological biases (Podsakoff et al.

2003). Furthermore, we explained to the participants that a robo-advisor is an application based on artificial intelligence that evaluates financial assets using a learning algorithm and analyzes historical data and current publicly available information. Based on this explanation, we surveyed our research models' constructs.

All main constructs were rated on a 7-point Likert scale, ranging from 'strongly disagree' to 'strongly agree' and can be found in Table 11 (Appendix). As by Krasnova et al. (2010), the items used to measure the effectiveness of the trust-increasing mechanism are self-developed since they had to be adapted to the specific conditions of robo-advisors and the comparison of mechanisms. The construct of relative advantage has been measured using the established items of Choudhury and Karahanna (2008), whereby all answers except 'strongly disagree' mean that an advantage is perceived up to a certain degree. We adopted the four factors of RA to our context of financial robo-advisory. Hence, RA-Learning describes the advantage of learning finance terms and concepts. RA-Informational Trust describes the extent to which trust exists in information obtained through a robo-advisor compared to the information obtained from a financial expert. RA-Informational Convenience describes how conveniently participants feel about receiving information via a robo-advisor compared to a financial expert. RA-Transaction describes the perceived convenience of the transactions and confidence in the transactions. We furthermore collected control variables.

5.4 Results

A total of 267 participants took part in the survey. In order to guarantee the quality of the answers, we have included an attention check. After removing samples that failed our attention check, 226 participants remained. Of these, 104 were female and 122 male. On average, the age of the respondents was 38.44, ranging from 18 to 68 years. Most of the participants were employed (61.1%), followed by students (13.3%). Hence, our sample is similar to the distribution of European Internet users (Eurostat 2018).

To test H1 we calculated the mean values and standard deviations of all four relative advantage constructs (see Table 6). Since the relative advantages of robo-advisors are perceived and the advantage decreases from RA-Informational Convenience to RA-Transaction, RA-Informational Trust and RA-Learning, H1 is supported. To find out if these differences of the relative advantage of using robo-advisors compared to human experts are significant, we used a one-way repeated measures ANOVA. A requirement for the use is that the variance of the difference between all relative advantage constructs is equal; this is called sphericity (Weinfurt

2002). Mauchly's test of sphericity indicated that the assumption of sphericity had been violated, $\chi 2(5) = 83.944$, $p = .000$. Epsilon (ε) was 0.794, as calculated according to Greenhouse and Geisser (1959), and was used to correct the one-way repeated measures ANOVA. The advantage was significantly different by the relative advantage constructs, $F(2.382, 536.032) = 10.190$, $p = .000$, partial $\eta 2 = .064$.

Construct	**Mean**	**SD**
RA-Learning (RAL)	3.717	1.417
RA-Informational Convenience (RAIC)	4.181	1.727
RA-Informational Trust (RAIT)	3.948	1.473
RA-Transaction (RAT)	3.951	1.555

Table 6. Mean and Standard Derivation of Relative Advantage Constructs

According to the post hoc test with Bonferroni adjustments all RAs are significantly different ($p < .001$) except for RA-Informational Trust and RA-Transaction ($p > 0.05$). The significantly preferred argument for the use of a robo-advisor is RA-Informational Convenience. RA-Learning is perceived as the weakest relative advantage and differs significantly from the other three RAs. RA-Informational Trust and RA-Transaction do not differ significantly from each other, leading to the following ranking results: 1) RA-Informational Convenience, 2) RA-Informational Trust, RA-Transaction, 3) RA-Learning.

We have also used a one-way repeated measures ANOVA to explore how effective various mechanisms to increase the users' trust are. The mean values and standard deviations of all eleven trust-increasing mechanisms are shown in Table 7. The mechanisms can be sorted as follows: *Testing, Data Transparency, Reasoning, History, Training Frequency, Confidence, Usage Time, Social Environment, Technical Functionality, Dialog, Visual Appearance*.

Mechanism	**Mean**	**SD**	**Rank**
Dialog (LOG)	3.93	1.716	3
Reasoning (RES)	4.67	1.678	2
Data Transparency (DAT)	4.73	1.619	2
Visual Appearance (VIS)	3.41	1.698	4
Confidence (CON)	4.56	1.743	2
History (HIS)	4.66	1.658	2
Testing (TST)	5.01	1.706	1
Social Environment (SOC)	4.50	1.784	2
Usage time (USE)	4.50	1.708	2
Technical Functionality (FNC)	4.43	1.675	2
Training Frequency (FRQ)	4.58	1.815	2

Table 7. Mean, Standard Derivation and Rank of Trust-Increasing Mechanisms

To test if reasoning increases trust more than the other mechanism (H2), we performed a one-way repeated measures ANOVA. As above, it must also be checked whether the assumption of sphericity is given. The assumption of sphericity had been violated as shown in Mauchly's test of sphericity, $\chi2(54) = 500.672$, $p = .000$. According to Greenhouse and Geisser (1959), we have calculated $\varepsilon = 0.610$ to correct the one-way repeated measures ANOVA. Trust growth was statistically significantly different between different measures, $F(6.105, 1373.535) = 44.649$, $p = .000$, partial $\eta2 = .166$.

The possibility of testing significantly increases trust in robo-advisors the most (all p-values are below 0.001), according to the post hoc comparison with Bonferroni adjustments. So we cannot confirm H2, which postulated that reasoning is the most trust-increasing mechanism. The visual appearance has the least significant influence on trust in robo-advisors, followed by talking in dialogue form (all p-values are below 0.001). All other mechanisms have no significantly different effect on trust in robo-advisors (p-values are above 0.05). However, there is a significant difference between knowledge about the database and knowledge about technical functionality, although yet these two mechanisms have no significant difference to the remaining mechanisms. These findings result in the following ranking: 1) Testing 2) Data Transparency, Reasoning, Confidence, History, Social Environment, Usage time, Training Frequency, Technical Functionality 3) Dialog 4) Visual Appearance. Therefore, we find that the mechanisms from the second rank are of equal importance in increasing the users' trust. However, we see a reason to presume that *Data Transparency* should be preferred over *Technical Functionality* because *Data Transparency* gains significantly more trust. We nevertheless included both mechanisms into the same rank since both mechanisms are of equivalent importance compared to all other mechanisms of ranking two.

5.5 Discussion and Implications

Due to technological developments, there are constantly more application possibilities for AI-based advisory systems as well as it is becoming more and more financially profitable to use AI-based advisory systems instead of human advisors. One of the challenges of using AI-based advisory systems is to gain user adoption and usage. To overcome this challenge, the question arises which advantages users see in AI-based advisory systems compared to classical human experts and which mechanisms can be used to increase trust in AI-based advisory systems. We know from the literature that the relative advantages are good for predicting adoption behavior (Choudhury and Karahanna 2008). In addition, many mechanisms have been discussed that increase the trust of (potential) customers in AI-based solutions (e.g., Ribeiro et al. 2016; de

Visser et al. 2016; Zanker 2012). Due to the scarcity of resources, in particular, companies cannot address all possible advantages and implement mechanisms to increase them, so it is important to know which advantages and which mechanisms have the greatest impact for users. We conducted a survey with 226 participants to answer our two research questions.

The results of the study have shown that users perceive a relative advantage in the usage of AI-based advisory systems (H1 is confirmed). As a consequence, the answer to the first research question is that users of expert systems currently perceive the superior capabilities of AI-based advisors compared to human advisors. This is an indicator that this innovation has a good chance of being adopted. Moreover, the most advantageous aspect of using AI-based advisory systems is convenient information retrieval, as expected for the informational stage. The objectivity and reliability of an AI-based advisory system are only perceived as a secondary advantage. The weakest perceived advantage is the possibility to learn something with the help of AI-based advisory systems. A possible explanation for this could be that the richness of the medium *robo-advisor* is not sufficient enough for the complex information in this context since it lacks the possibility to ask questions or discuss specific features.

We were able to establish a ranking from the results of the study to determine the preferred mechanisms for increasing trust in AI-based advisory systems. This ranking shows that most trust is gained when users are able to use and test the system without risk (H2 is not supported). This also corresponds to the results of Zuboff (1988), according to which trust in new technologies depends to a large extent on the possibility of trying them out. This is also reflected in Rogers (2003) view, which assumes that there are three dimensions: experience, understandability, and observability. Trust based on understanding the machine is also much more stable than trust based on performance (Lee and See 2004). The answer to our second research question is that the highest level of trust can be generated with the help of the mechanism testing.

Anthropomorphism as mechanism had the weakest impact on trust. This finding can be explained with the literature on automation bias, which states that the initial level of trust is higher in automated systems (Dzindolet et al. 2003). Nevertheless, such human features could have an impact on regaining trust when the robo-advisor was performing badly (de Visser et al. 2016). The mechanisms *experience*, *subjective norms* as well as *transparency* had almost the same impact on trust. However, within the group of *transparency* mechanisms, providing information about the used *data*, gains more trust in the AI-based advisory system than providing information about the *technical system* that is used for the robo-advisor.

The **theoretical contribution** of our findings lies in complementing the JAS and innovation adoption literature in various ways. Although the JAS literature considers trust to be the largest antecedent for advice adoption (Van Swol and Sniezek 2005), our results show that this is not the case for AI-based advisory systems. By replacing a human with a machine, a big change occurs. Such a change is also called innovation (Rogers 2003). In innovation adoption research, it is shown that the RA-Informational Convenience, and not the trust in innovation is decisive to recognize the relative advantage of an innovation. It follows that not only the JAS literature alone is sufficient for the impact and response to an AI-based advisory system, but that it must be considered in combination with the innovation adoption literature. In addition, we were able to demonstrate that the impact of different mechanisms presented in the literature on trust in AI-based systems varies. Therefore, future research should not only consider whether one mechanism has an influence, but how effective it is compared to others.

Furthermore, our results have **practical implications** in addition to the theoretical contribution. AI-based advisory systems can be described as innovations that are still in their infancy. For companies intending to offer such products, the question arises whether this innovation has chances of success and how these can be increased. Our results show that people see relative advantages in the use of AI in advisory systems. Therefore, companies can be confident when introducing an AI-based advisor that it will be adopted. In addition, companies can become aware of the advantages users see in AI-based applications compared to human experts. However, there is a potential upward in the perception of relative advantages. AI-based advisory firms can increase trust in their product or service using a variety of mechanisms. However, it is not possible for a company to implement every possible mechanism to increase their customers‘ trust. The implementation of each mechanism is financially expensive and the needed resources (e.g., human resources) are intensive. Furthermore, some mechanisms (e.g., providing explanations) might not be technical realizable or only by accepting losses in performance, yet. Based on our study results, companies can prioritize the implementation of the mechanisms to increase trust. In this way, those mechanisms can be implemented that have the greatest impact on users and are therefore the most advantageous. Therefore, companies should first offer to test their products and in the next step start to make the system more transparent, start with transparency regarding the data used.

5.6 Limitations, Future Research and Conclusion

However, there are also some limitations to our study. In order to measure the benefits of using AI-based advisory systems and to measure which mechanisms increase trust, we have chosen robo-advisor as the optimal context. Despite the fact that the use of scenarios in IS Advice research is a common method (e.g., Jung 2018; Wang and Benbasat 2007), it makes it difficult to generalize our findings. The rankings could be different in a different context. In addition, the mechanisms for increasing trust in AI-based advisory systems were specifically selected for robo-advisors and the items were self-developed, so it would be important to validate these results in a different context as well as in an experiment. Moreover, it has been shown that RA-Informational Convenience is the strongest argument for the adoption of AI-based advisory systems. In the next step, it would be interesting which mechanisms exist to influence this advantage.

Technological development is making AI-based systems more relevant. AI-based systems can be used in various ways. One possibility is to use them as advisors, e.g., as financial advisors. However, in order to benefit from these AI-based advisory systems, they must be adopted and used by users. To the best of our knowledge, it was not investigated if users perceive an advantage of AI-based advisors and only isolated trust-increasing mechanism were considered in the previous IS literature (Hegel et al. 2009; Nilashi et al. 2016). Our study shows that the biggest advantage that people see in the usage of such AI-based systems is easy access and convenient use. In addition, trust in such systems can be increased faster through the opportunity to test the system than through other mechanisms. With the help of these results, companies that have scarce resources can better prioritize their decision to market AI-based systems and use mechanisms that gain trust in their products or services

6 Paper D: Investigating Users’ Utilization of Advice from Robo-Advisors

Title

Following the Robot? Investigating Users’ Utilization of Advice from Robo-Advisors

Authors

Tauchert, Christoph, Technische Universität Darmstadt, Germany

Mesbah, Neda, Technische Universität Darmstadt, Germany

Publication Outlet

Proceedings of the 40th International Conference on Information Systems (ICIS 2019), Munich, Germany

Abstract

Companies are gradually creating new services such as robo-advisors (RA). However, little is known if users actually follow RA advice, how much the fit of RA to task requirements influences the utilization, how users perceive RA characteristics and if the perceived advisor’s expertise is influenced by the user’s expertise. Drawing on judge-advisor systems (JAS) and task-technology fit (TTF), we conducted an experimental study to measure actual advice-taking behavior in the context of RA. While the perceived advisor’s expertise is the most influential factor on task-advisor fit for RA and human advisors, integrity is a significant factor only for human advisors. However, for RA the user’s perception of the ability to make decisions efficiently is significant. In our study, users followed RA more than human advisors. Overall, our study connects JAS and TTF to predict advice utilization and supports companies in service promotion.

Keywords

artificial intelligence; robo-advisor; advice taking; judge-advisor system; task-technology fit; user's expertise

6.1 Introduction

Current advances in artificial intelligence (AI) are driving companies to develop new services for their customers. Giving an example, in the light of Industry 4.0, manufacturing firms offer their clients the possibility of predictive maintenance or process optimization based on machine data (Rawal 2019). Besides these newly offered services, enterprises are also transforming established services by empowering them using machine learning and make them scalable by taking the human out of the loop. An example of these kinds of changes can be seen in the traditional service sector, such as the legal or financial industry. Typically, legal or financial advisory is done by experts who advise you on how you should act with regard to your specific needs. In recent years, empowered by AI, companies have developed services that give personalized advice using an information system instead of a human expert (HSBC 2017).

In the literature, there is no common definition for AI. Russel and Norvig (2009) define AI based on two dimensions. The first dimension addresses the thought process and behavior, while the second one is concerned with whether success is measured against human performance or against an ideal (rational) performance. The combination of these two dimensions leads to four characteristics which can describe and define an AI-based system: thinking humanly, thinking rationally, acting humanly, or acting rationally (Russel and Norvig 2009). In our context of AI-based advisory, we define AI as a system, which is able to learn, makes rational predictions, and interacts like a human. AI differs significantly from other traditional technologies since AI-based systems do not just follow predefined static rules but have the ability to learn from data (Burrell 2016). Some advantages of AI-based systems are efficiency and scalability (Brundage et al. 2018). In comparison to human advisors, AI-based advisors are not able to explain their recommendation, which is also known as black-box behavior, but due to technological advances, AI-based algorithms can process, utilize, and learn from more information than any human advisor could do in appropriate time because of cognitive constraints (Simon 1972).

A much-discussed example of AI-based advisory, in research and practice, is financial robo-advisory which causes significant changes in the financial industry (Jung, Dorner, Glaser, et al. 2018; Jung and Weinhardt 2018; Sironi 2016). Robo-advisors are automated investment advisory services. Customers are guided through a self-assessment process and are then recommended a target-oriented investment strategy with regard to possible portfolio compositions or estimated stock performances (Jung, Dorner, Glaser, et al. 2018; Sironi 2016;

Tertilt and Scholz 2017). If robo-advisors were accepted by users[1], benefits would arise both for providers as well as for users. Due to the simple scalability of advisory services as well as the significant reduction of investment costs, the deployment of robo-advisors is highly attractive for financial service companies like banks (Tertilt and Scholz 2017). By using robo-advisors, users can also reduce their investment costs and perform real-time portfolio surveillance (Tertilt and Scholz 2017).

Assuming robo-advisors can provide good advice, it is not guaranteed that people will necessarily utilize such advice. In the information systems (IS) literature robo-advisors were mostly investigated focusing on the design and architecture of these services as well as related business models (Eickhoff et al. 2017; Jung, Dorner, Weinhardt, et al. 2018; Jung and Weinhardt 2018; Riasanow et al. 2018). Whereas, the exploration of users' perception of robo-advisors was neglected. In the cognitive sciences, the judge-advisor system (JAS) paradigm has been used to investigate the advice taking and giving behavior of people. Although various factors were examined within this research stream, almost exclusively the interaction between human decision makers and human advisors was regarded (Bonaccio and Dalal 2006). In the IS literature, the task-technology fit (TTF) is used to determine how well a technology is suited to assist a person in performing a task (Goodhue and Thompson 1995a). However, based on this model, we cannot assess if AI-based advice is utilized differently than human advice. By integrating the TTF model in the JAS context, we want to generate a holistic view to understand the factors leading to advice accepting behavior of AI-based advisory services. Therefore, we examine if users accept a substitution of human financial advisors by robo-advisors and if the investment advice will be at least similarly utilized. This leads us to the following research questions:

RQ 1: *Are there differences in users' advice utilization of robo- and human advice?*

RQ 2: *Is the users' advice utilization affected by the fit of task and advisor as well as how this fit is affected by the advisor's characteristics?*

Since the topic of finance and financial planning concerns the general population (Beketov et al. 2018), it is natural that both experienced and inexperienced individuals might use robo-advisors. Users' perceived expertise was already discussed within the TTF as well as the JAS literature (Harvey and Fischer 1997; Parkes 2013). JAS researchers have shown that experienced decision makers have higher advice utilization when making important decisions

[1] User is defined as user of a robo-advisor and used synonymously to decision maker and judge.

(Harvey and Fischer 1997). Furthermore, Parkes (2013) found that users' expertise affects the perception of technology characteristics. Therefore, we explore if users' self-perceived expertise has an impact on the perceived expertise of human and robo-advisors. Consequently, our third research question is:

RQ 3: *Does the users' self-perceived expertise affect the perceived advisor's expertise?*

We are following the call of Rzepka and Berger (2018) to investigate users' interactions with robo-advisors by answering these three research questions. The remainder of this manuscript is structured as follows: To begin with, we provide an overview of the theoretical background related to advice taking and the task-technology fit. Then, we derive hypotheses before describing our online experimental survey study design. After introducing our study sample consisting of 197 participants, we present the collected and analyzed data using group comparison and partial least square (PLS). Thereby, the discussion of findings illustrates contributions to research and practice. Lastly, we conclude the manuscript by summarizing the most important findings as well as pointing out the limitations of our research and proposing specific avenues for future research.

6.2 Theoretical Background

6.2.1 Advice Giving and Taking

Within the cognitive sciences, the phenomenon of people giving and taking advice is investigated under the judge-advisor system (JAS) paradigm (Bonaccio and Dalal 2006). It describes a structured group in which one individual (i.e., the judge or decision maker) holds the sole decision power and seeks advice from one or more advisors (Van Swol 2011). Within this context, various studies have investigated which factors influence the judge's advice utilization, i.e., the extent to which decision makers follow the advice they receive from experts (Bonaccio and Dalal 2006). A robust finding has been *egocentric discounting*, which means that decision makers tend to adjust their initial estimate by just 20% to 30% towards the advisor's suggestion (Harvey and Fischer 1997). In addition, several factors such as trust, competence, distance of advice, power or source of advice have been identified as influencing the advice-taking behavior of decision makers (Bonaccio and Dalal 2006; Schultze et al. 2015; Sniezek and Buckley 1995; Van Swol and Sniezek 2005; White 2005).

One of the most discussed advisor characteristics influencing advice-taking is trust (Jungermann 1999; Van Swol 2011). Trust is "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (Mayer et al. 1995, p. 712). Since trust is a rather abstract concept, most researchers agree that it has to be studied multi-dimensionally (Komiak and Benbasat 2006; Rousseau et al. 1998). Komiak and Benbasat (2006) categorized trust in three dimensions: (1) cognitive trust in competence, (2) cognitive trust in integrity, and (3) emotional trust. Following their definition, several studies analyzed the impact of an **advisor's competence**, also called **expertise**, on the decision maker's advice utilization. Advisor's competence is defined as the advisor's perceived ability to provide good advice in a specific domain (Mayer et al. 1995). Customers' main concern is whether the advisor has the competence required to provide them with relevant and customized advice (Komiak and Benbasat 2006). Studies have shown that decision maker, who perceive their advisor as competent are more willing to adjust their initial opinion in favor of the advisor's opinion. (Kim et al. 2017; Schultze et al. 2015). **Integrity** is defined as the honesty of the advisor and describes the decision maker's expectation that the advisor acts in his/her interest (McKnight et al. 2002). Consequently, it refers to the extent that the user perceives the advice as objective and unbiased (Komiak and Benbasat 2006). A robo-advisor can be designed in a way that it only recommends products that are most profitable for the service provider who owns the robo-advisor. Such kind of robo-advisor would be considered to have a low integrity. Studies show that the higher the perceived integrity of the advisor is, the more likely it is that the advice will be used (Van Swol 2011). Lastly, **emotional trust** describes the decision maker's feelings of security and comfort about relying on an advisor (Komiak and Benbasat 2006). Similar to the previous dimensions of trust, the stronger the emotional trust in the advisor is, the more likely the judge is to follow the advice (Sniezek and Van Swol 2001).

Besides the advisor's expertise, also the decision maker's expertise was investigated. While generally it is assumed, that the decision maker has a lower competence than the advisor, decision makers with less expertise had higher trust levels and more variability in their trust rating (Sniezek and Van Swol 2001). It has been shown that experienced users tend to follow advice less than inexperienced users (Harvey and Fischer 1997). For incompetent decision makers it is difficult to assess whether a particular advice is good or bad (Ehrlinger et al. 2008), which is a huge challenge in the judge-advisor relationship.

Concluding, in the JAS literature, advisor's and decision maker's expertise, advisor's integrity as well as emotional trust in the advisor have been identified to influence advice utilization. Based on the task-technology fit model, we want to combine these factors. Therefore, we will introduce the TTF model in the next section.

6.2.2 Task Technology Fit

For a technology to be adopted by the user it has to be utilized and it must help the user to achieve his/her goal in a specific task. A well-known theory used for this phenomenon is the task-technology fit (TTF) (Goodhue 1995; Goodhue and Thompson 1995a).

The task-technology fit model consists of the components *task*, *technology*, the *fit*, and the *utilization of the information system*. A task describes any action that is carried out to turn inputs into outputs. Relevant task characteristics are those that influence individuals to use or not to use a technology (e.g., difficulty, significance, routineness). A technology is a tool that an individual uses to carry out a task. The fit describes the appropriateness in which a technology helps the individual to succeed in a task. This fit should serve as a predictor for the utilization of an information system since individuals are more likely to use a technology that they perceive to be suitable to assist in solving the task (Goodhue and Thompson 1995a). Figure 7 depicts the general idea of the model.

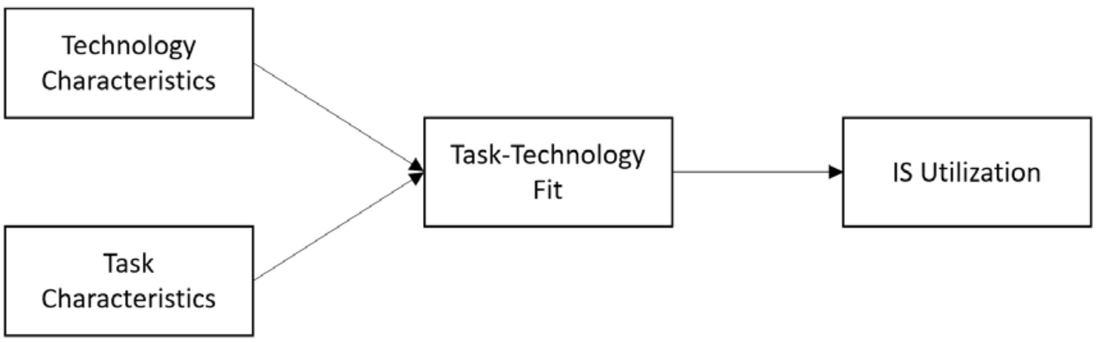

Figure 7. Task-Technology Fit Model (adapted from Goodhue 1995)

The TTF was already used in various contexts to investigate the success of new technologies including answering managerial questions (Goodhue et al. 2000), online shopping (Klopping and Mckinney 2004), question-answering systems (Robles-Flores and Roussinov 2012) and group support systems (Zigurs et al. 1999; Zigurs and Buckland 1998). However, until now it was not used to evaluate the setting of robo-advisory systems.

6.3 Research Model

The purpose of this manuscript is to investigate and compare the behavior of individuals when interacting with robo-advisors and human experts in a financial planning context using the judge-advisor system. Until now, the JAS paradigm was almost exclusively used in a setting where both the judge and the advisor were human beings. However, there is one study which investigated how individuals utilize advice that is deducted from a statistical model (Önkal et al. 2009). Although the advice was presented in the exact same way for the statistical method and the human advisor, the participants discounted the statistical advice more than the same advice from a human expert. While robo-advisors are also mostly based on statistical methods, they have more capabilities. As some studies have shown, they might be perceived differently since human characteristics are perceived in AI-based applications (Rzepka and Berger 2018). Furthermore, compared to human experts, AI algorithms are able to process a vast amount of information in real-time and can incorporate the resulting insights in their advice (Anthes 2017). This implies that robo-advisors must be seen as more than purely statistical tools and this could lead to an increased reliance on robo-advisors due to a perceived superiority:

H1: *Advice from robo-advisors is utilized more than advice from human experts.*

The TTF describes the fit between task characteristics and a technology. By adapting this to the JAS context, the task-advisor fit (TAF) describes the fit between task characteristics and an advisor. Since TTF is a predictor for IS utilization (Goodhue and Thompson 1995a), we assumed that TAF would be a predictor for advice utilization:

H2: *A higher task-advisor fit is related to higher advice utilization.*

From the JAS literature, we know that trust is identified as one of the most important factors that lead to advice utilization. Other characteristics such as age (Feng and MacGeorge 2006) and similarity to the decision maker (Gino et al. 2009) were also investigated. Many of these factors are not directly transferable to robo-advisors. Therefore, we focused on the advisor characteristics that can be perceived in a human advisor as well as in a robo-advisor.

To validate the advisor characteristics that were identified through the literature review, we conducted a pre-test among 67 persons. We asked the participants (1) what characteristics they see in a human advisor, (2) what characteristics they associate with a robo-advisor and (3) what differences between those two types of advisors they perceive. The open answers were coded by three IS researchers and as a result, we can confirm the literature-based characteristics but also found that efficiency-enhancing was an often mentioned characteristic, that describes the extent to which an advisor enables efficient decision-making. Therefore, we considered four advisor characteristics: expertise, emotional trust, integrity, and efficiency-enhancing.

As mentioned before, studies have shown that decision maker, who perceive their advisor as competent are more willing to adjust their initial opinion in favor of the advisor's opinion. (Kim et al. 2017; Schultze et al. 2015). Furthermore, advisors with higher expertise are able to assess the difficulties and challenges of a task better and thus, are more suitable to solve a task successfully. Therefore, we hypothesized:

H3a: *For the robo-advisor, a higher perceived advisor expertise is related to higher task-advisor fit.*

H3b: *For the human advisor, a higher perceived advisor expertise is related to higher task-advisor fit.*

A great advantage of robo-advisors is their ubiquity since they are available for consultation 24/7 other than human financial advisors. Furthermore, they provide advice instantaneously because of their superior data processing capabilities. Therefore, robo-advisors enable users to make investment decisions more efficiently. In the case of the human advisor, efficiency will not be a decisive factor when it comes to whether the advisor is perceived as suitable. Nonetheless, due to the access to the advisor's additional expertise, efficiency in decision-making increases. Thus, leading to the following hypotheses:

H4a*: For the robo-advisor, a higher perceived advisor efficiency-enhancing ability is related to higher task-advisor fit.*

H4b: *For the human advisor, a higher perceived advisor efficiency-enhancing ability is related to higher task-advisor fit.*

From the JAS literature, we know that trust has a major influence on advice utilization (Jungermann 1999; Van Swol 2011). Emotional trust describes the feeling of security and comfort about relying on the advisor (Komiak and Benbasat 2006). Thus, the decision maker perceives the advisor as credible and helpful, leading to a positive influence on TAF:

H5a: *For the robo-advisor, a higher emotional trust is related to higher task-advisor fit.*

H5b: *For the human advisor, a higher emotional trust is related to higher task-advisor fit.*

When interacting with an advisor, decision makers cannot be sure of the advisor's intentions. It is not necessarily clear, whether the advisor is advising in the best interest of the client or if he/she acts for their own personal gains. Especially in the context of financial advice this topic gained some media coverage with advisors maximizing their commission fees and kickbacks. Therefore, decision makers will deem the advisor suitable for the task if they perceive them to have a higher integrity:

H6a*: For the robo-advisor, a higher perceived advisor integrity is related to a higher task-advisor fit.*

***H6b**: For the human advisor, a higher perceived advisor integrity is related to a higher task-advisor fit.*

As described before, a main problem of incompetent decision makers is to evaluate whether the received advice is correct and useful (Ehrlinger et al. 2008). Therefore, only if a user has a certain task knowledge, he/she can assess the expertise of an advisor. Thus, we concluded:

***H7a:** For the robo-advisor, a higher self-perceived user expertise is related to higher perceived advisor expertise.*

***H7b:** For the human advisor, a higher self-perceived user expertise is related to higher perceived advisor expertise.*

Since the focus of our study lies in the perceptual differences of robo- and human advisors, we assessed the model using only one task and did not manipulate any task characteristics. Nonetheless, we measured a set of task characteristics such as significance as well as difficulty (Petter et al. 2013) and we did not find any variation in the task characteristics. Finally, Figure 8 shows the final research model, which builds the foundation of our study.

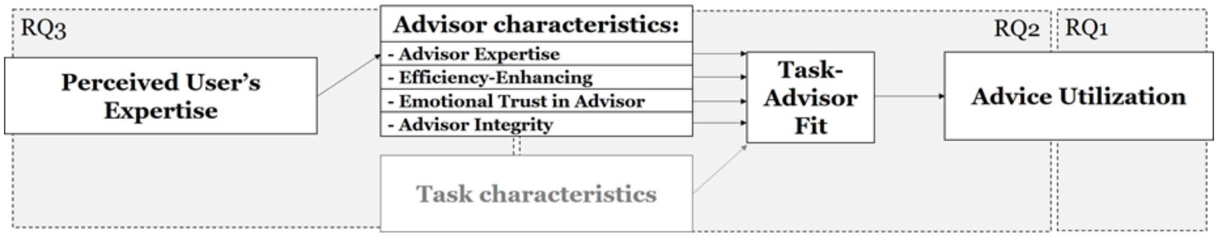

Figure 8. Research Model

6.4 Research Method

In order to investigate the differences between the utilization of advice from a robo-advisory system and a human expert, we set up an online experimental survey. Our goal was to measure the participants' actual behavior during the interaction with the advisor instead of their self-reported perception as it was called upon by Rzepka and Berger (2018). We developed an experiment following the approach of many studies in the JAS context (e.g., Gino and Moore 2007). The participants were randomly assigned into two groups whereby one group was instructed that their advice comes from a human expert and the other group was told that the advice is given by a robo-advisor. In order to motivate the participants to reveal their true intentions, they had the chance to win up to 2 Euro during the experiment if they perform well (Camerer and Hogarth 1999).

To acquire a diverse and highly representative (in terms of age, gender, and occupation) sample of internet users for our study, we used a market research company (Lowry et al. 2016). The participants received an incentive of 0.5 Euro from the agency regardless of their performance during the experiment. At the landing page of our study, the participants were instructed that they participate in a scientific study, that their data is stored anonymously and, that, besides the experiment task, there are no right or wrong answers. This was done to counter common methodological biases (Podsakoff et al. 2003).

6.4.1 Experiment Description

To answer the research questions and validate our hypotheses, we decided the context of stock prediction to be a good fit for our survey. While this approach does not reflect the typical interaction process of clients and robo-advisors, it allowed us to apply a widely recognized measure within the JAS research stream (i.e., the weight of advice). Furthermore, this use case deemed to be appropriate for four reasons: (1) Users are familiar with stock prices due to daily news coverage. (2) The prediction of stock prices is a part of robo-advisory services since it is necessary to recommend a good stock portfolio. (3) The prediction of stock prices is not just a knowledge-based task due to the high uncertainty of stock markets (Dzielinski 2012). (4) Finally, it is also very important that advice is reliable as it has a long-term negative effect in the event of failure for users (Lee 2009).

The study was structured as followed: At first, we collected the participants' demographics before having them answering some self-assessment constructs. Then, a description of the experiment scenario was shown to the participants: „Imagine: You want to invest in company shares and must, therefore, forecast the performance of various stocks. Your task is to estimate how a particular stock will perform within a year." Furthermore, they got the information that they will see real historical stock valuation charts from the recent past and that the closer their final estimation is to the real stock valuation the higher their compensation will be. The experiment roughly consisted of three repeating steps per stock:

1. Analyzing the provided stock chart and giving an independent initial estimation.
2. Getting the valuation estimation of an expert, which was either a human or robo-advisor.
3. After receiving the expert's opinion, the participants were free to adjust their initial estimation. They were explicitly told that they could but do not need to change their personal estimation.

After the scenario description, the experiment began and the participants were sequentially shown five charts showing a 3-year historical (*t* to *t+3*) stock performance of enterprises out of five different industries (i.e., aviation, pharmaceutical, automotive, technology, and energy). Additionally, the participants received a small description (one sentence) about the enterprise. We withheld the information of the exact timeframe and the companies' names to avoid that individual experiences were weigh in that might distort advice utilization (Önkal et al. 2009). After each chart, the participants had to guess the stock valuation a year later (*t+4*). After they had estimated the last stock value, the participants were told that they now get professional advice from an expert. The first group was informed that the expert is a professional (human) financial advisor, who had a profound education in finance and is founding advice on his/her experience, current news, and economic developments. The second group was told the advice comes from a robo-advisor, an application based on AI that uses historical stock data, analyzes current news as well as economic developments to generate an advice. The provided advice was the same for both groups and corresponded to the true stock valuation. Afterward, the participants were again shown the charts sequentially with the additional information of their initial estimation as well as the advisor's estimation. The participants were asked to give a final estimation of the expected stock valuation.

6.4.2 Items

To measure the degree of advice utilization we used the *weight of advice* (WOA), which has been used in several studies (e.g., Gino and Moore 2007; Önkal et al. 2009; Sah et al. 2013; Schultze et al. 2015):

WOA = (final estimate – initial estimate) / (advice – initial estimate)

The WOA measures to what extent an individual utilizes an advice in his final estimation by dividing the distance of final and initial estimate by the distance of advice and initial estimate (Yaniv 2004). For rational decision makers the WOA is supposed to be in the range of 0 and 1. 0 meaning that the participant completely ignored the advice and did not adjust his/her initial estimate and 1 implicating that the decision maker completely adopted the advice. Values in-between 0 and 1 indicate partial incorporation of the advice in the final estimate, whereby a value of 0.5 means that a participant has calculated the mean between his/her initial estimate and the advice and weighs his/her opinion just as much as the advisor's. Irrational decision makers can have WOA measure under 0 or over 1, meaning that either moved in the opposite direction of the advice or that he/she even over-utilized the advice. However, these cases occur rarely (Gino and Moore 2007; Harvey and Fischer 1997). We calculated the mean WOA using the five measured WOA values for each participant.

For the evaluation of the constructs, we have used measurements from the established literature. We used the scales of Komiak and Benbasat (2006) to measure emotional trust and cognitive trust in integrity. To measure trust in competence we adapted the scale of McKnight et al. (2002). The perceived efficiency-enhancing ability of the advisor was measured using the construct of Chan et al. (1997), while we used Moore and Benbasat's (1991) scale for the task-advisor fit. Finally, we adopted the item of Radel et al. (2011) to measure user's self-perceived task expertise. All of our items were measured using a 7-point Likert scale ranging from 'strongly disagree' to 'strongly agree' and can be found in the appendix in Table 12. Additionally to the items of our main constructs, we measured tendency towards fantasizing as marker variable to counteract common method bias (Podsakoff et al. 2003) based on the three-item scale of Darrat et al. (2016).

6.5 Results

In our study 247 participants took part. We included several checks – manipulation check and rationality check – to guarantee the quality of the study's results (Meade and Craig 2012). During the rationality check, we excluded all participants who had a WOA over 1 or under 0. We excluded 21 participants due to failing the manipulation check. After excluding 29 more participants who failed the rationality check, our sample consisted of 197 responses, which could be used for further analysis. The demography of our sample reflects the typical European internet users quite accurately by age, gender, and employment status (Eurostat 2018). 93 females and 104 males took part in our study with an average age of 38.54 years ranging from 18 to 68 years. 58.4% of our participants were employees and 11.2% students. From our remaining participants, 104 were assigned to the group with the robo-advisor while 93 participants were assigned to the group with the human advisor. In order to compare the behavior of both groups, we first ensured that the groups had perceived task characteristics equally and that user's self-perceived expertise was not significantly different by using an independent t-test.

H1 hypothesized that advice from the robo-advisor would be more utilized. To test H1, we ran an independent t-test of WOA. The result of the t-test ($t(195) = 1.771$, $p = .039$) showed that the advice of the robo-advisor was statistically more utilized ($M = .44$, $SD = .253$) than those of a human advisor ($M = .38$, $SD = .260$). Concluding, H1 is supported.

To evaluate H2 to H7, we analyzed our research model based on a well-establish method (Qureshi and Compeau 2009) by comparing the structural equation model of each group through a variance-based partial least squares multi-group analysis as implemented in SmartPLS (Ringle et al. 2015). We opted for this approach for two main reasons. (1) This approach is well suited for theories in their early stages (Fornell and Bookstein 1982). (2) It is possible to test both the research models and the path differences simultaneously through multi-group analysis (Brook et al. 1995).

Constructs (measured on 7-point scales)	**Items**	**Item Loadings Robo-Advisor**	**Item Loadings Human Advisor**
Advice Utilization (WOA)	WOA1	1.000	1.000
Task-Advisor Fit (TAF)	TAF1	.861	.919
	TAF2	.952	.936
	TAF3	.936	.913
Advisor Expertise (AEX)	AEX1	.928	.944
	AEX2	.961	.960
	AEX3	.930	.957
	AEX4	.872	.927
Advisor Efficiency-Enhancing (EFF)	EFF	1.000	1.000
Emotional Trust in Advisor (EMO)	EMO1	.961	.970
	EMO2	.977	.971
	EMO3	.973	.977
Advisor Integrity (INT)	INT1	.906	.881
	INT2	.931	.916
	INT3	.919	.926
User's Self-Perceived Expertise (UEX)	UEX1	.941	.937
	UEX2	.975	.961
	UEX3	.959	.959
	UEX4	.938	.929

Table 8. Item Loadings

By determining convergent validity (statistical similarity of construct items) and discriminant validity (statistical difference of items that measure different constructs) of our research model we validated our measurement model (Hair et al. 2013). We confirmed convergent validity by examining item loadings, Cronbach's α, and composite reliability (CR) as well as the average variance extracted (AVE) by the constructs (Xu et al. 2012). The item loadings are reported in Table 8 and all loadings are above the threshold value of 0.7 (Hair et al. 2013). For each construct the Cronbach's α and composite reliability values achieve the threshold of 0.7 and AVE values threshold of 0.5 (Hair et al. 2011) as can be seen in Table 9.

We assessed the cross loadings as well as the square root of the AVE for each construct model and therefore, we confirmed discriminant validity (Fornell and Larcker 1981). As reported in

Table 9, all constructs' square roots of the AVE are higher than their correlation to another construct. The loading of each item is greater to its associated construct than to other constructs, but we do not report the cross loadings due to space limitations.

Constructs	**Cr. α**	**CR**	**AVE**	**WOA**	**TAF**	**ACOM**	**EFF**	**EMO**	**INT**	**SCOM**
Advice	1.000	1.000	1.000	**1.000**						
Utilization (WOA)	1.000	1.000	1.000	**1.000**						
Task-Advisor Fit	.905	.941	.841	.233	**.917**					
(TAF)	.913	.945	.851	.300	**.923**					
Advisor Expertise	.942	.958	.852	.396	.789	**.923**				
(AEX)	.962	.972	.897	.334	.783	**.947**				
Advisor Efficiency-	1.000	1.000	1.000	.173	.724	.723	**1.000**			
Enhancing (EFF)	1.000	1.000	1.000	.255	.737	.768	**1.000**			
Emo. Trust in Advisor	.969	.980	.941	.421	.686	.790	.724	**.970**		
(EMO)	.971	.981	.946	.435	.663	.703	.711	**.973**		
Advisor Integrity (INT)	.908	.942	.844	.288	.583	.669	.568	.647	**.918**	
	.894	.934	.824	.280	.737	.672	.738	.794	**.908**	
User's Expertise	.967	.976	.909	-.146	.167	.181	.220	.100	.171	**.953**
(UEX)	.962	.972	.896	.015	.130	.172	.102	.187	.145	**.947**

Table 9. Cronbach's α (Cr. α), Composite Reliability (CR), Average Variance Extracted (AVE) and Construct Correlations (First Row: Robo-Advisor; Second Row: Human Advisor)

Before we test our research model through the multi group analysis, we depict the results of the research model for the full sample by running a bootstrapping with 5,000 re-samples (Davison and Hinkley 1997). As we postulated in H2 higher task-advisor fit relates significantly to a higher advice utilization (ß = .273, p = .000). H2 is supported. We can also find a significant impact of each advisor characteristic on task-advisor fit – for advisor expertise (ß = .480, p = .000), for advisor efficiency-enhancing ability (ß = .253, p = .004) and for advisor integrity (ß = .141, p = .050) – except for emotional trust (ß = .036, p = .732). User's self-perceived expertise has a significant impact on perceived advisor expertise (ß = .175, p = .026). None of our control variables – age, gender, IT background, marker variable for common method bias – changes the significances of our research model or are significant predictors of our dependent variable.

At the beginning of the multi-group analysis, we looked at the model fit of both groups. The model fit SRMR is .056 for the robo-advisor sample and for the human advisor sample .052, which refers to a good model fit since it is under the cut-off value of .08 (Hu and Bentler 1999). Based on our model we are able to explain 5.4% of the variance of the advice utilization and 67.4% of the variance in task-advisor fit in the robo-advisor sample and 9.0% of the variance of the advice utilization and 70.4% of the variance in task-advisor fit in the human advisor sample. The path coefficients, their significance as well as their effect sizes are reported in Table 10.

Likewise to the full sample research model, H2 is supported in the multi-group analysis. As postulated in H3a and H3b, advisor expertise has a significant impact on task-advisor fit. The advisor's efficiency-enhancing ability has a positive significant influence on task-advisor fit for robo-advisors, as assumed in H4a. However, we have to reject H4b since there was no significant influence of the efficiency-enhancing ability on task-advisor fit for human advisors. Contrary to our assumption of H5a and H5b, emotional trust has no significant influence on task-advisor fit for either group. Although advisor's integrity has a significant positive influence on task-advisor fit in the group with the human advisors as postulated in H6b, it has no significant influence on task-advisor fit in the robo-advisor group against our suggestion of H6a. Finally, we have not observed a significant effect of the user's self-perceived task expertise on advisor expertise. Summarizing the results, we were able to support H2, H3a, H3b, H4a and H6b. All other hypotheses had to be rejected. These findings of our multi-group analysis are visualized in Figure 9.

Constructs	**Path Coefficients and p-Values**				**Multi-Group Testing**		**f² values**	
	RA	**p**	**HU**	**p**	**Diff.**	**p**	**RA**	**HU**
Task-Advisor Fit → Advice Utilization	.233*	.017	.300*	.011	.067	.673	.057	.099
Advisor Expertise → Task-Advisor Fit	.516***	.000	.461***	.000	.055	.376	.244	.258
Advisor Efficiency-Enhancing → Task-Advisor Fit	.304**	.006	.165	.228	.140	.211	.117	.029
Emotional Trust in Advisor → Task-Advisor Fit	.028	.825	-.057	.726	.085	.343	.001	.003
Advisor Integrity → Task-Advisor Fit	.047	.562	.351**	.002	.305*	.017	.003	.128
User's Expertise → Advisor Expertise	.181	.079	.172	.191	.009	.487	.034	.030

Table 10. Results of Structural Model Testing and Effect Sizes (* p<0.001; ** p<0.01; * p<0.05; RA=Robo-Advisor, HU=Human Advisor)**

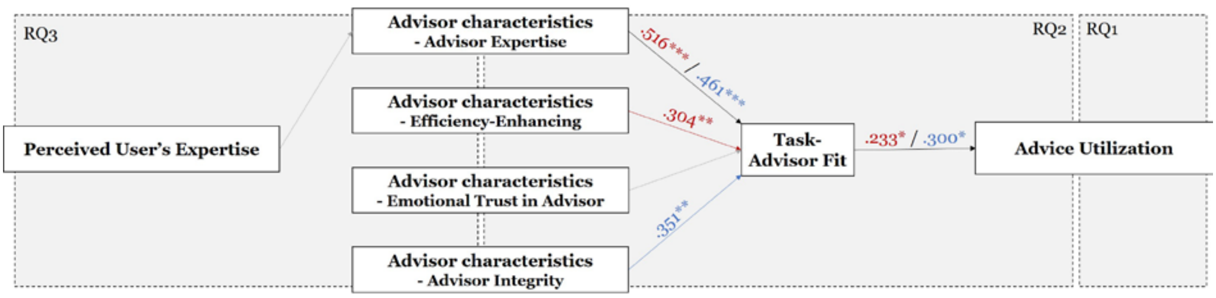

Figure 9. Summary of Structural Model Testing (Robo-Advisor in Red and Human Advisors in Blue)

6.6 Discussion and Contribution

The goal of our research was to investigate (1) whether users utilize advice differently depending on the source of advice (i.e., robo-advisor vs. human expert), (2) if the task-advisor fit affects advice utilization as well as how advisor characteristics influence the task-advisor fit and (3) the influence of users' self-perceived expertise on the perception of the advisor's expertise. To address our research questions, we conducted an experimental study with 197 participants and thereby contributed to the IS advice-taking literature.

Previous studies have shown that the origin of advice can have a significant influence on the user's utilization of advice. It has been shown that advice that is derived from statistical models is discounted more than advice from human experts in a financial setting (Önkal et al. 2009). Other studies that have investigated the perception of 'traditional' computer-generated advice have also found that human advice is trusted more (Wærn and Ramberg 1996). Our experiment's findings showed that the advice of robo-advisors was utilized more than the advice of human experts for the specific setting of stock price predictions. To understand the differences of the findings in our study, we argue that while robo-advisors base their advice mostly on statistical and mathematical calculations one can interact with robo-advisors more naturally due to natural language processing and speech synthesis abilities. Therefore, the advantages of both advisor types are combined. However, this result needs to be validated in further studies and causalities have to be derived.

With regard to the advisor characteristics, we found that in our context different characteristics affect the task-advisor fit for the different advisors. For the robo-advisor, we can see that expertise and efficiency-enhancement are the significant antecedents while for the human advisor, expertise and integrity are contributing to the task-advisor fit. Even though we had to reject H6a (i.e., a positive influence of integrity on TAF) when calculating an independent t-test, we noticed that for robo-advisors a significantly higher integrity is perceived than for human advisors ($m_{AI} = 5.12$, $m_{HU} = 4.09$, $t(195) = 5.74$, $p = .000$). This could be an indication that users suspect the dishonesty of humans but do not believe that robo-advisory services give malicious advice. Nonetheless, by comparing the f^2 values it can be seen that for both cases expertise was the most influential antecedent. Concluding, human and robo-advisors are perceived with different strengths that are influencing the task-advisor fit. Furthermore, it can be seen that the TAF is a predictor for advice utilization and the integration of the TTF model in the JAS paradigm was successful.

Another finding of our study is the influence of the decision maker's self-perceived expertise on the perception of the advisor's expertise. Even though we hypothesized a positive relationship between these two constructs, we do not find a significant effect. This is quite surprising because the decision makers' knowledge in the area of interest should allow them to assess the quality of the given advice better (Ehrlinger et al. 2008). Since our advisor always gave the true estimation, we expected that the competent users would rate the advisor's expertise higher. However, they did not have much information about the advisor and the interaction was different in comparison to a real consultation. Therefore, it might have been difficult to evaluate the advisor's expertise. Additionally, we did not measure the participants' real expertise but rather the self-perceived expertise. This perception could be overestimated. The Dunning-Kruger effect describes that incompetent individuals often do not know that they are incompetent (Kruger and Dunning 1999). So this effect could lead to a false self-assessment of our participants and consequently an overestimation of their expertise.

Besides the theoretical contributions, we also identified various practical implications for professional entities: we found first evidence that AI-based financial advice could be utilized more than human advice by users. This indicates that enterprises can deploy robo-advisors without generally having to fear that customers will reject the suggestions. Furthermore, since the task-advisor fit might be a predictor for the actual advice utilization, organizations can conduct market research surveys to assess the suitability of potential robo-advisory services. Enterprises can leverage our findings about robo-advisor characteristics to adjust service development. They could increase the perceived robo-advisor's expertise, for example, by providing more transparency about the used data or the algorithm so that the assessment of the workflow and performance would be easier. Another option is providing key performance indicators, which enable simpler evaluation of the (historical) performance. Lastly, organizations could emphasize the efficiency of robo-advisors for personal financial planning.

6.7 Limitations and Future Research

Naturally, the findings of our study are subject to various limitations. First of all, we selected the setting of robo-advisors as context and used a stock valuation experiment to measure advice utilization. While scenario-based experiments are a common method in IS research (e.g., Önkal et al. 2009; Wang and Benbasat 2007; Ye and Johnson 1995), the findings need to be validated in other robo-advisor tasks such as portfolio composition, especially since stock evaluation is not a typical task during the interaction of financial advisors with customers. Furthermore, since robo-advisors can be used in a variety of domains such as in the legal or insurance industry, one could select a different experimental setting. The investigation of tasks that have different task characteristics (e.g., difficulty, significance, locus of control, (non-)routines) could be very interesting. To give an example, task difficulty has been found to have an impact on advice utilization as well as self-perceived expertise (Ehrlinger et al. 2008; Gino and Moore 2007). If it is possible to predict the perfect advisor characteristics based on the task, promising use cases for AI-based advisors could easily be identified.

Furthermore, we compared the perception of robo-advisors and human ones based on an online experiment. We assumed that participants could put themselves in the situation of a real consultation with a human financial advisor by describing the scenario. It would certainly be useful to validate the findings of our study in a more realistic laboratory experiment where participants would interact with a real human advisor and robo-advisor. The authentic interaction with a human and robo-advisor could lead to different perceptions of advisor characteristics like emotional trust, expertise, or integrity.

There are various other different experimental designs that could also be considered in future works: For example, we did not offer the option to choose between two advisors. It could be interesting to investigate the behavior of users when they have a choice between different advisors. Additionally, our experiment required the advisor to provide a numerical estimation, but there are plenty of other types of advice (e.g., advice for sth., advice against sth., binary advice) that can be studied. Furthermore, user expertise was the sole individual's characteristic that was within the scope of our study. Certainly, various other personality traits may influence the perception of advisor characteristics (e.g., confidence, introversion).

To summarize this section, our task-advisor fit model is a first approach to integrate the TTF model in the JAS to understand users' perceptions of robo-advisors and to evaluate the resulting advice utilization.

6.8 Conclusion

Due to current technological developments and advancements in the area of artificial intelligence, AI-based agents are gaining importance in enterprise services rapidly. Such agents can be implemented in a wide variety of fields such as in the healthcare, legal or as in our case the financial industry. The use of robo-advisors is currently gaining momentum, but market shares of such services are still relatively low (Jung and Weinhardt 2018). Therefore, the goal of this manuscript was to investigate users' utilization of advice from robo-advisors. In addition, we wanted to explore if the users' advice utilization is affected by the fit of task and advisor as well as how this fit is affected by the advisor's characteristics. Furthermore, the influence of the user's self-perceived task expertise on the perception of the advisor's expertise was addressed.

By conducting a scenario-based experimental study with 197 participants in a European country, placed in the context of financial advisory, and using performance-based incentives, we were able to measure actual advice utilization. Thus, we were able to show that: (1) Users utilize advice from a robo-advisor differently than advice from a human expert. In our setting the users utilized the advice from robo-advisors more than the advice from human advisors. (2) Users perceive different advisor characteristics for robo- and human advisors. In our experimental setting for the robo-advisor, competence and efficiency were perceived as characteristics that influence the task-advisor fit and for human experts, the significant factors were competence and integrity. (3) The user's self-perceived task expertise has no influence on the perception of the advisor's expertise. Our results help to understand the factors influencing how robo-advisor services are perceived by users and what drives them to utilize the advice from these services. Based on our findings, companies can focus on relevant factors when designing and implementing a robo-advisory service.

7 Paper E: Towards an Integrative Approach for Automated Literature Reviews Using Machine Learning

Title

Towards an Integrative Approach for Automated Literature Reviews Using Machine Learning

Authors

Tauchert, Christoph, Technische Universität Darmstadt, Germany

Bender, Marco, Technische Universität München, Germany

Mesbah, Neda, Technische Universität Darmstadt, Germany

Buxmann, Peter, Technische Universität Darmstadt, Germany

Publication Outlet

Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS-53), Weilea, Hawaii, USA

Abstract

Due to a huge amount of scientific publications which are mostly stored as unstructured data, complexity and workload of the fundamental process of literature reviews increase constantly. Based on previous literature, we develop an artifact which partially automates the literature review process from collecting articles up to their evaluation. This artifact uses a custom crawler, the word2vec algorithm, LDA topic modeling, rapid automatic keyword extraction, and agglomerative hierarchical clustering to enable the automatic acquisition, processing, and clustering of relevant literature and subsequent graphical presentation of the results using illustrations such as dendrograms. Moreover, the artifact provides information on which topics each cluster addresses and which keywords they contain. We evaluate our artifact based on an exemplary set of 308 publications. Our findings indicate that the developed artifact delivers better results than known previous approaches and can be a helpful tool to support researchers in conducting literature reviews.

Keywords

text analytics, LDA, literature review, text mining

7.1 Introduction

Due to the advancing digitization, more and more data is being generated in a wide variety of areas, including science (Bornmann and Mutz 2015). For example, in mid-2019 over 560,000 documents are found in all EbscoHost databases for the keyword search "artificial intelligence" (AI). The number of scientific publications is increasing immensely. Although these papers are mostly accessible, the information is prevalently unstructured (i.e., available as PDF file) (Nair and Narayanan 2012). A fundamental task of researchers is to discover and understand the existing literature through a literature review in order to establish the context and conduct new and further research (Dann et al. 2017). For this purpose, it is essential that all existing literature relating to a research topic is reviewed. However, this task is hardly feasible with the constantly increasing number of papers and their evaluation is practically difficult. To cope with the huge amount of publications, researchers might be supported by an IT artifact for structured literature reviews, which collects available documents and provides first insights of the existing literature.

Recent developments in technology, especially in machine learning, enabled (partially) automated literature reviews to become technically feasible. AI is a sub-field of computer science containing techniques such as machine learning, deep learning and natural language processing to enable intelligent machines (Elliot and Andrews 2017; McCarthy 2007). AI is efficient and scalable (Brundage et al. 2018) and provides capabilities to enable a machine to process more information and gain deeper insights than any human being can because of their cognitive constraints (Simon 1972). In the past, several attempts to use data mining to solve specialized problems similar to automated literature reviews (e.g., medical case analysis (Huang et al. 2005)) have been made. However, there is still no well-established method how this new technology can be used to perform a (partially) automated literature review. We therefore try to address the research question: How can an IT artifact be designed to support researchers in conducting structured literature reviews?

To our best knowledge, only Dann et al. (2017) developed an artifact that uses the word2vec algorithm and keyword extraction to automate the literature review process based on full-text papers. Nevertheless, their approach has weaknesses. For example, each paper still has to be downloaded manually, which is a challenge with these large quantities of papers. Additionally, identifying the theme of a cluster is not easy and still involves a lot of work.

The aim of our research is to extend their approach, so that the whole process from collecting the data, processing and evaluating the clusters becomes simpler and more reliable.

To achieve this goal, we first sum up at the related literature, where we focus in particular on the approach of Dann et al. (2017). Then we describe the design science method on which we have based and further developed the artifact. Afterwards, we present and evaluate our artifact. Finally, we summarize our results.

7.2 Related Research

Literature reviews play a crucial role in research and science since the creation of new knowledge is often based on the interpretation, combination and questioning of already existing knowledge (Schryen et al. 2015). However, conducting a literature review is very time consuming and cumbersome due to the many manual activities, such as searching and downloading, documentation of the process, text screening, etc. Nonetheless, knowing and understanding the results and findings of existing literature is crucial to contribute to research and helps to avoid investigating what has already been investigated (Schryen et al. 2015).

One of the most renowned and widely-used process models in IS research for conducting a (manual) literature review is the framework of vom Brocke et al. (2009), which is depicted in Figure 10. This framework is often used in conjunction with the concept matrix suggested by Webster and Watson (2002). This matrix helps to understand and link the various concepts used in the processed publications.

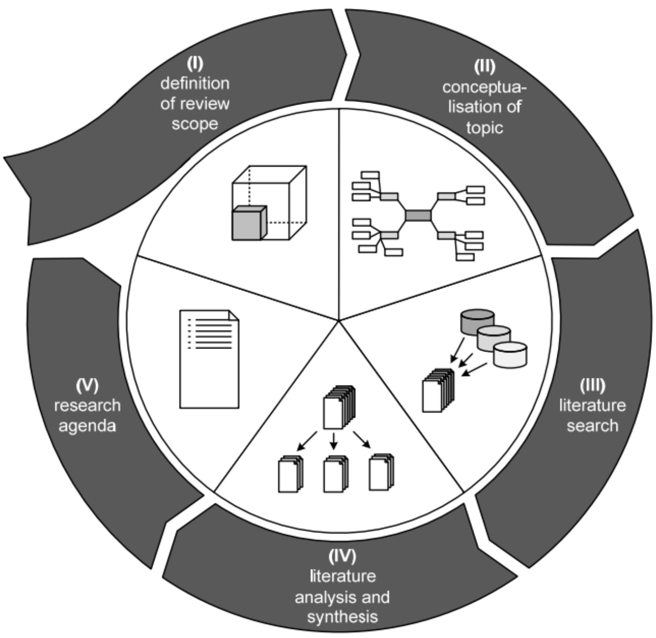

Figure 10. Framework for Literature Reviews According to vom Brocke et al. (2009)

Regarding the usage of algorithms to analyze scientific documents, Dann et al. identified three categories: (1) citation-based approaches which only consider the links between documents by analyzing the references, (2) text-based approaches which analyze the actual textual context of the documents and (3) hybrid approaches which combine the two former mentioned approaches (Dann et al. 2017).

Since we focus on a content-based analysis to extract knowledge from existing literature, we are only considering manuscripts that deal with text-based approaches. Furthermore, text-based approaches are considered superior to citation-based ones for document categorization (Aljaber et al. 2010). The used approaches differ in three aspects: (1) text sections (i.e., abstract, keywords, full text), (2) objective (e.g., classification, recommendation, content extraction, clustering), and (3) used techniques (e.g., bag-of-words, vectorization, Bayesian classifier, topic models, keyword extraction) (Afonso and Duque 2014; Dann et al. 2017; Gulo et al. 2015; Wang and Blei 2011).

While we were inspired by and based our artifact on Dann et al.'s (2017) presented process, we suggest an extension of their model to improve information extraction and automation. By implementing a crawler to download all documents related to a search term automatically, a very effortful but rather trivial task is automated. Furthermore, by implementing a topic model using Latent Dirichlet Allocation (LDA) we add another analytical layer to gain more insights into the gathered literature. The extracted topics can especially be helpful to identify common concepts of the scientific publications supporting the derivation of Webster and Watson's concept matrix.

7.3 Design Science Research

The proposed solution is an IT artifact in accordance to Hevner et al. (2004). Due to the fact that the proposed solution solves a problem that is primarily targeted and based on the "*science business*", it can be categorized as an idiographic design science artifact since it is an "ideal artifact for a specific problem" (Baskerville et al. 2015). Identifying and analyzing a manifold of databases, searching relevant literature and analyzing content is on the one hand a very time-consuming task for scientists but a partly structured task with some repeating actions on the other hand which makes it suitable for automation – not just in the science context (e.g., (Günther and Quandt 2015)). Developing a solution, which partially supports or entirely replaces this part of the research work is therefore highly relevant for the scientific community. The growing number of publications in journals or conference proceedings as well as other potentially less scientific sources poses a great challenge to researchers since a comprehensive and extensive overview of a certain topic becomes harder to attain (Aljaber et al. 2010; Bornmann and Mutz 2015; Huang et al. 2005). Data-mining techniques have the potential to overcome these challenges and especially text-mining can be applied to the unstructured data that scientific publications usually contain (Aljaber et al. 2010; Dumouchel and Demaine 2006). Current research and knowledge discovery processes generally have a very low degree of automation and are vastly done by humans instead of algorithms since these tasks are usually less structured compared to manufacturing processes, for example.

Therefore, the support of a software-based, semi-automated knowledge extraction tool has several (practical) benefits for the researcher and can overcome the described restrictions:

- *time*: software-based solutions can gather and analyze publications faster than humans. Furthermore, the process can easily be parallelized and is therefore scalable.
- *structure*: the research process is always executed identically. Also, every publication is analyzed in the same way and results are processed alike. This guarantees repeatability and reduces subjectivity and personal bias in the evaluation process (Lacity and Janson 1994).
- *cross-disciplinarily*: although the way research is conducted varies across several science areas, the preparation of a research project is in most cases at least similar. This also means that the automated process can be transferred across disciplines and is ideally not restricted to a specific, singular discipline, like Information Systems (IS) research. Positive effects of this standardization and automation are increased comparability and eventually improvements in the generation of insights and explanations can be achieved in a standardized way (Martens and Provost 2014).

The overall implementation is similar to the process described by Dann et al. (2017). An extension our solution provides is the addition of the actual information retrieval process. Relevant publications are identified by using the search interfaces of online databases and full texts as well as bibliographic information are downloaded automatically. This not only speeds up the whole research preparation phase, it also enables further filtering after the contents are locally available and document selection/filtering is no longer restricted to the capabilities of the database and does not require manual inspection and evaluation of whether a document fits the required search criteria or not. The Python programming language was chosen to implement the software artifacts since it provides platform independence and many readily available packages that are common in the data science process. The following steps of data preparation and processing are implemented analogously to established text-mining processes: conversion of full-text PDF files to machine-readable text files, vectorization of text files, clustering, and keyword extraction (e.g., Dann et al. 2017). At the end of the process, the results are visualized and presented to the user.

The artifact was developed in an iterative manner. Initial requirements were successively extended and the solution were implemented and tested according to the additional requirements. It therefore is designed as a search process in which an initial solution was continuously enhanced and refined to reflect the process of preparation steps to knowledge extraction better step by step (Hevner et al. 2004).

Since there is little guidance in the IS literature on how to evaluate design-science research (Pries-Heje et al. 2008), our approach for the evaluation of the artifact is achieved in a three-stage process. First, the artifact was used in a specific context, i.e. the actual functionality of the data acquisition, filtering, pre-processing and clustering was evaluated. Second, the results of the first steps were evaluated by humans who determined whether the proposed clusters are correct and useful. Lastly, the clusters themselves were discussed by a group of four IS researchers (Hevner et al. 2004).

7.4 ALR Approach

Our artifact contains of the following steps: (1) downloading documents and making documents machine-readable, (2) preprocessing downloaded full- text documents, (3) vectorising documents, (4) extracting keywords, (5) identifying topics, (6) clustering documents and (7) visualization (see Figure 11).

As many data science projects are written in Python, our artifact likewise is also largely implemented in Python. The data acquisition is separated in several classes since it is a more complex task. The process was then orchestrated in Jupyter notebooks, which are a convenient way to combine code and documentation. At this point, efficiency and performance were not the main goals, instead, the focus lied on building an easily readable and reusable code base with an understandable interface. Jupyter notebooks also allow for rapid code changes and integrating visualization.

7.4.1 Collecting and generating machine-readable documents

The first element of our artifact has the purpose of downloading the documents that will be analyzed in the subsequent process steps. Therefore, we wrote a web crawler for two established scientific databases (i.e., EbscoHost and ScienceDirect), that downloads all available publications that match the user-defined *search terms* and *date range*. While ScienceDirect offers an application programming interface (API) to search and download plain text documents, we used the Python library *Selenium* to download documents from EbscoHost. This library allows to "remote control" a web browser to navigate the EbscoHost website and extract the required information without an API.

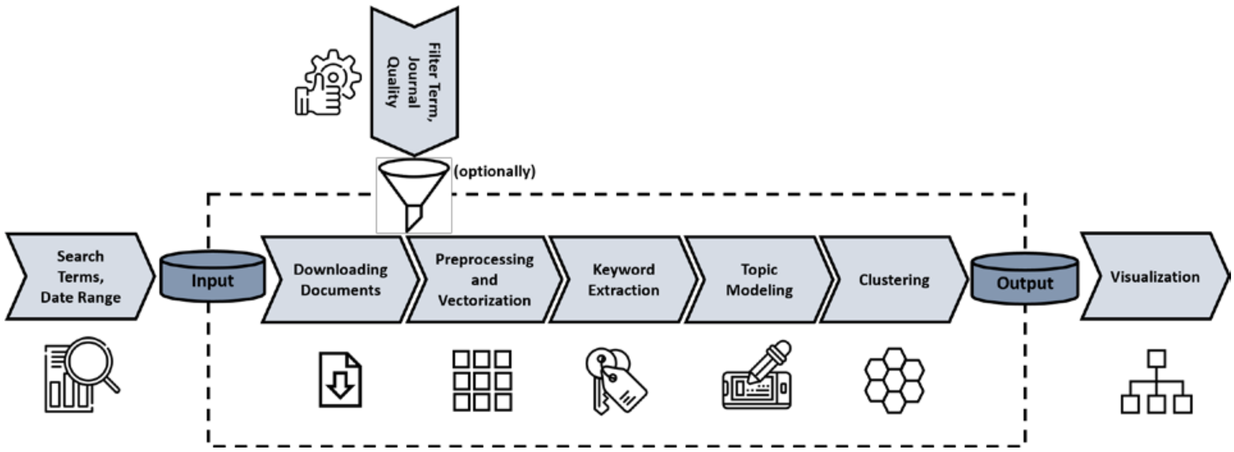

Figure 11. Automatic Literature Review Process

In contrast to the ScienceDirect API, documents that were downloaded through EbscoHost were only available as PDF files and were therefore not directly machine-readable. Older manuscripts are often embedded images of scanned manuscripts, which cannot be directly processed by a text-mining algorithm. Similarly, newer files usually contain the plain text version of the publication but due to the proprietary binary format it is difficult to extract the text using open source libraries. To overcome these issues, we used the open source optical character recognition (OCR) library *Tesseract* that can convert the PDF files into plain text. Tesseract is a popular and widespread software for OCR and is currently developed by Google, whose engineers utilize it themselves for text recognition on mobile devices, for example. The output of Tesseract is a plain text file comparable to the files that were retrieved via the ScienceDirect API.

The result of this step is a collection of plain text documents of all scientific documents available through EbscoHost and ScienceDirect. Furthermore, the program uses a database to store meta data about the retrieved documents such as title, authors, year, journal, etc.

7.4.2 Text preprocessing

With all relevant documents available in a machine-readable plain text format, the next phase of the process can be conducted: the preprocessing of the textual data.

Depending on the search criteria (i.e., search terms and date range) specified in step 1, the automated download of documents may lead to a large number of documents, which the user might want to restrict in retrospect. For that reason, we added the possibility to reduce the set of documents using *filter terms*. All documents, which do not contain the *filter terms*, are then excluded from further processing. Therefore, filter terms can also be used to change the scope of the analysis iteratively based on the insights gained through previous analysis, e.g., extracted keywords, topics or frequent words. The effectiveness of reducing the document set by using filter terms depends largely on the specificity of the filter terms and the heterogeneity of the documents. Additionally, we implemented the option to consider only documents from journals that have a Q1 score in the Scimago Journal Ranking. This enables the user to limit the analysis to results from journals with a certain quality. Both filtering features are optional and the users can decide whether they want to exclude documents based on filter terms and/or journal ranking.

The preparation of documents for analysis usually contains a text-cleaning step. Hereby, punctuation and stop words (e.g., *for, and, of,* etc.) as well as user-defined words are removed from the text corpuses. In our case, this list contained words such as journal names, placeholders for figures, etc. Furthermore, words are normalized to their word stem (e.g., *fisher, fishing, fishy* are reduced to *fish*). The result of this step is cleaned and stemmed textual data.

7.4.3 Vectorization of documents

To apply a hierarchical clustering approach to the corpus of the collected documents, their text needs to be represented by vectors of a fixed length. One of the most applied models to transform the representation of documents into vectors is the bag-of-words model. Despite its simplicity and efficiency, it often achieves a high accuracy. Texts are represented as unsorted collection of the contained words. The model then assigns weights to the words, which represent the frequency in the document and in the collection of documents. Documents with similar word frequencies can be considered as having similar contents (George and Joseph 2014). The simplicity of the model also yields some major drawbacks. The model does not consider the order of words, which leads to the problem that different semantics of sentences with different order cannot be distinguished. Furthermore, ambiguity and synonyms cannot be considered. Ambiguity means that words can have different meanings depending on the surrounding context. Synonyms are words that are different but do have the same meaning.

Due to the aforementioned drawbacks of the bag-of-words model, we generate the vector representation by using the paragraph-vector model which is based on the word2vec model (Le and Mikolov 2014; Mikolov et al. 2013). These models take the context of words into account (i.e., the paragraph) and therefore partly solve the issues of the bag-of-words model. The paragraph-vector algorithm learns continuously distributed vector representations of texts of any length by using artificial neural networks to learn a word vector for any word in the document collection (Ai et al. 2016; Le and Mikolov 2014). Each document can therefore be represented as a structured concatenation of word vectors.

The calculated word vector representation for the documents allows to compare similarities of documents by using common distance measures. We calculate a document **X** document distance matrix, which can then be used for hierarchical clustering.

7.4.4 Keyword extraction

Keywords are often used to tag documents for the purpose of information retrieval (Rose et al. 2010). There are many available approaches to automatically generate keywords, which can be categorized in statistical, supervised, semi-supervised and unsupervised approaches (Siddiqi and Sharan 2015). In our context, only statistical and unsupervised methods can be used, since the downloaded documents are not labeled.

An algorithm for keyword extraction that is often used in practice is RAKE (Rapid Automatic Keyword Extraction) (Rose et al. 2010).

RAKE assumes that keywords frequently consist of multiple words but rarely contain standard punctuation or stop words (i.e., *and*, *the*, and *of)*, or other words with minimal lexical meaning. Therefore, the algorithm uses these stop words and phrase delimiters to create a list of candidate keywords by partitioning the text at these positions. Afterwards, the graph of co-occurrences is computed and word scores are calculated. The word score is the ratio of word degree and word frequency. The word degree is the sum of co-occurrences and favors words that occur often and in longer candidate keywords. The word frequency is the pure number of occurrences of a word in the candidate list. Due to the partition of the text using stop words and phrase delimiters, candidate keywords cannot contain any stop words, such as in *illusion of control*. To tackle this problem the algorithm then searches for pairs of keywords and creates a combination of two keyword candidates if they adjoin one another at least twice in the same document. Afterwards the *N* top scoring candidates are selected. A proposed number of keywords is one-third the number of words in the graph of co-occurrences (Mihalcea and Tarau 2004).

7.4.5 Topic identification

To gain more insights on the clustered documents, we added another analytical layer to the process. By implementing a probabilistic topic modeling approach, we can get an overview of predominant topics within the clusters. In our case, we use LDA as topic model. LDA assumes that documents cover multiple topics, which can be seen as a distribution over a defined vocabulary (Blei et al. 2003). Furthermore, it is assumed that topics existed even before the documents were written. LDA tries to invert the imaginary random process and therefore guesses which hidden topic structure has probably generated the observed document collection (Blei et al. 2010, 2003).

To use LDA, we have to specify the number of topics that should be identified. There are several evaluation metrics to assess the appropriateness of a topic model (Arun et al. 2010; Cao et al. 2009; Deveaud et al. 2014; Griffiths and Steyvers 2004). A typical approach is to calculate multiple metrics and determine the number of topics by aggregating the provided information. Figure 12 exemplary shows the result of the calculation of the four metrics for 10 to 450 topics for the Associated Press data set. While the metric by Deveaud et al. (2014) is not informative in this case, the other three metrics reach their minima or maxima in the area of 90 to 140 topics.

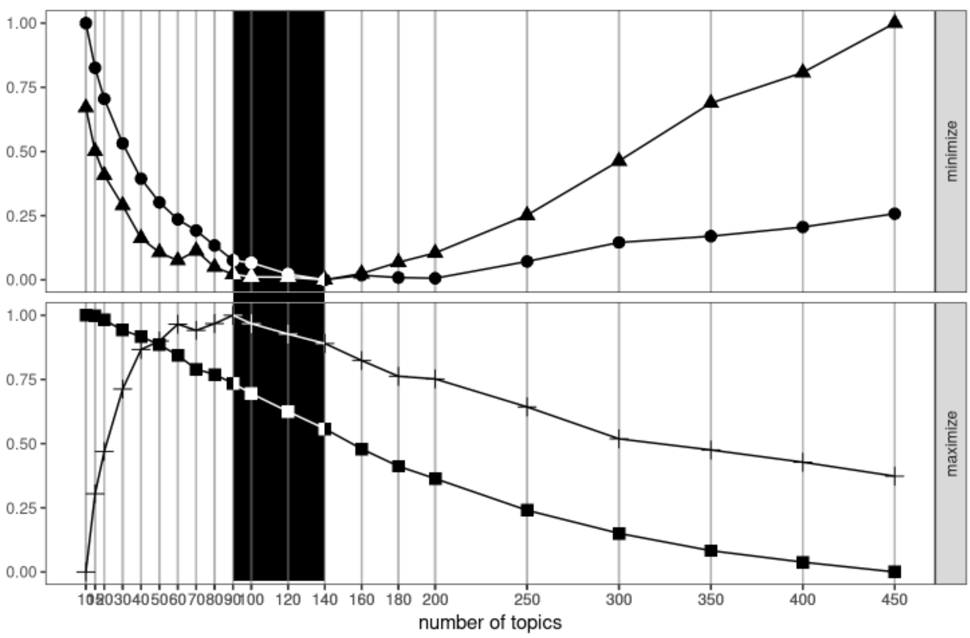

Figure 12. Selecting the Number of Topics (Nikita 2019)

The result of this step is a topic model, a list of identified topics and their related words as well as the information which topics are present in each document.

7.4.6 Clustering of documents

Clustering is a very popular approach when it comes to text mining and its goal is to form groups of similar documents by detecting hidden patterns within them. All documents contained in one cluster should be similar to the other documents in the same cluster but different to documents contained in every other cluster (Berkhin 2006).

One popular clustering algorithms is *k-means clustering*, which is widely used in data mining. This algorithm creates *k* clusters, with *k* being a number that has to be specified beforehand. It then maximizes the sum of squared deviations between documents in different clusters (Kriegel et al. 2017).

While there are multiple methods to estimate the optimal number of clusters, such as the elbow criteria (Thorndike 1953) and silhouette score (Rousseeuw 1987), we did not get useful results in our context. The elbow criteria usually suggested for form two clusters while the silhouette score preferred as much clusters as possible.

Therefore, we decided to use a more flexible approach without the need to specify the number of clusters beforehand: agglomerative hierarchical clustering. Hereby, the algorithm calculates a tree-like hierarchy, which can easily be visualized using a dendrogram and enables an explorative data analysis with varying granularity. Agglomerative clustering initially creates a cluster for every object and recursively joins those clusters until a cutoff-value for the distance between the clusters is reached (Madhulatha 2012).

7.4.7 Visualization of analysis

Since the set of documents usually is rather large, depending on the chosen search and filter terms, we chose adapt the representation of result. While there is also information provided on the entire data set, the more detailed information such as keywords and topics are only provided within the generated clusters. This approach counter-acts information overload (Burkhard 2004).

The generally provided information consists of an overview of (1) the distribution of documents across clusters, (2) which journals are represented, (3) the distribution of publishing dates and (4) the identified topics.

The information per cluster consists of: (1) number of documents in cluster, (2) number of identified LDA topics, (3) range of publishing dates, (4) distribution of LDA topics, (5), extracted keywords, (6) most common words and two-word phrases, (7) represented journals in the cluster, (8) dendrogram with titles, (9) dendrogram with LDA topics, (10) dendrogram with keywords, (11) dendrogram with authors.

7.5 Evaluation

For the initial evaluation of our approach, we decided to use a rather small set of documents. This allows us to assess the quality of generated keywords, topics and clusters manually and properly. Although the document crawler was working as intended, we used an existing set of 308 documents on the *application and usage of sensor data in industrial manufacturing* due to the occurrence of licensing issues and resulting time restrictions.

Figure 13 shows the distribution of documents across the 17 generated clusters with cluster sizes ranging from nine to 29. Due to the used approach of agglomerative hierarchical clustering, the created partition is not the "one ground truth" but one of many possible partitions. Other partitions might be more specific or more general.

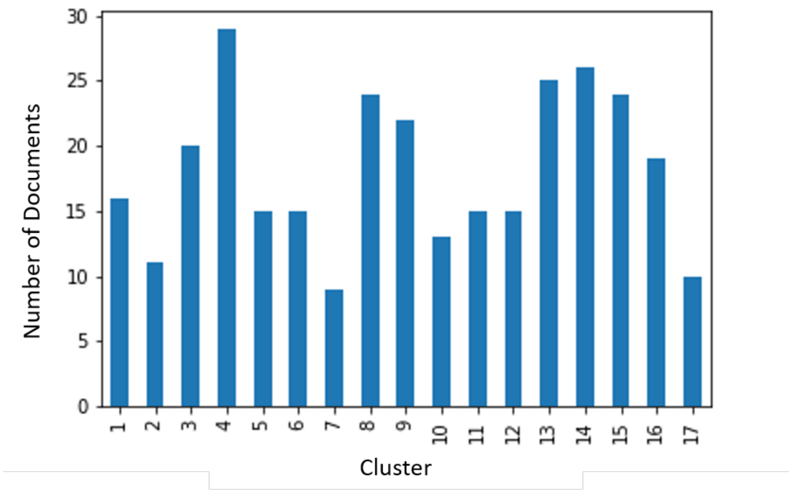

Figure 13. Distribution of Documents across Clusters

To get an overview of the clusters and therefore the similarity of the documents, a dendrogram depicts the distances between the documents and clusters (see Figure 16 in Appendix A3). We added titles for each cluster, which we derived by inspecting the algorithmically extracted information (i.e., keywords, the most common words and topics). For the LDA topic model, we decided to use 33 topics since the evaluation metrics mentioned in section 4.5 suggested that 25 to 40 topics would be appropriate. Exemplarily we show three of the found topics and the words they consist of:

1) Real-time RFID technology: "real_time", "rfid", "shop_floor", "material", "production" "task", "technology", "location", "product" "operator"

2) Intelligent Grinding (with Industry 4.0): “grinding”, “industryfourzero”, “production”, “level”, “wheel”, “rule”, “expert”, “intelligent”, “grinding_wheel”, “rowe”

3) Real-time Fault Detection and Simulation in Assembly: “real_time”, “error”, “event”, “simulation”, “assembly”, “station”, “degradation”, “line”, “exception”, “service”

The remainder of this section describes the information that is visualized (not computed) for every isolated cluster to prevent information overload. Due to space restrictions, we are describing and evaluating the data only for the first cluster *Process Monitoring and Error Diagnostic for Assembly Stations.*

Figure 14 shows the most prevalent topics for the first cluster. The higher the percentage share of the top topics, the more homogenous the cluster is. When all topics have a relatively low percentage share, the cluster is more diverse. We can see that the first cluster is about error detection at assembly stations, tonnage signals and detecting errors on wafers.

Cluster 1: Average LDA Topic Distribution
29.1% Errors at assembly stations
28.7% Error diagnostic with tonnage signals
25.2% Error diagnostic and recognition at wafers
7.1% Algorithms for error and quality classification
3.1% Detecting surface errors with machine vision

Figure 14. LDA Topics of Cluster 1

This information can be enriched by also considering the automatically extracted keywords. The keywords are partitioned by number of words and sorted in declining order of their RAKE-score (see Figure 15). Therefore, important keywords are listed first. The number of documents within the cluster that contain this keyword is shown in brackets.

Cluster 1: Keyword (number of documents containing the keyword)
state space model(6)
manufacturing process(10), process monitoring(6), principal component(10), assembly process(11), assembly station(8), fault diagnosis(11), fixture layout(9), dimensional quality(7), sensor location(8), locating scheme(4), fixture fault(9), autobody assembly(7), variation pattern(7), pattern recognition(9), engineering knowledge(6), standard deviation(7), sample size(9), dimensional variation(9), measurement data(8), proposed method(12), proposed methodology(5), covariance matrix(8), final product(7), measurement point(7), locating pin(5), degree freedom(5), assumed independent(4)
table(16), distance(7), monitoring(10), line(16), time(13), contribution(9), coordinate(9), application(12), sensor(14), quality(13), based(16), change(14), process(15), limit(13), set(16), component(14), variable(15), design(14), assembly(11), control(13), analysis(14), developed(13), diagnosis(14), position(12), source(10), station(12), performance(12), structure(11), model(15)

Figure 15. Extracted Keywords of Cluster 1

Using the extracted keywords, it can also be seen that this cluster is about monitoring assembly processes mainly in the context of car bodies. Apparently, in many documents principal component analysis is used. The placement of sensors and pattern recognition seem to be of importance.

By looking at the dendrogram (see Figure 17 in Appendix A3), we can easily understand that the first three documents deal with process monitoring and diagnosis for stamping or forging while the two following documents tackle the problem of thermal errors at machine tools. The analysis of the dendrogram of keywords (not included due to space limitations) shows that these documents were matched because the same approach (i.e., principal component analysis) was used. The documents in the red frame are more homogenous and deal with error diagnosis at assembly station in manufacturing processes. The topic dendrogram of the cluster (see Figure 18 in Appendix A3) reveals an anomaly within the cluster, since the most important topic of the sixth paper does not appear in the other documents of the cluster.

Summarizing the evaluation, we conclude that the proposed artifact leads to useful results for clustering the selected publications in the context of industrial manufacturing and therefore supported the subsequent analysis and synthesis of the literature effectively. The topics contained within the clusters were described well by the extracted keywords. The extension of introducing an additional topic model has proven useful in further understanding of the extracted topics and differentiation of inter-cluster homogeneity and heterogeneity of topics. The topic model also helped by splitting single clusters into multiple sub-clusters. The cognitive load that is put upon the researcher if clustering is done manually was (subjectively) reduced significantly by using this approach compared to a solely manual literature review of the same number of publications. Manual structuring, synthesizing and describing of these manuscripts would also have required significantly more time. Nonetheless, the task of creating a literature review does not become trivial by just introducing a (partially) automated solution. The researcher is still required to understand and process the literature and to extract relevant knowledge. Especially a close inspection of the decision rules for assigning a publication to a topic is necessary, since they might not always reflect the researchers own expectations. For example, a cluster which is determined by association of authors could be less suited than a cluster that is chosen because of contextual proximity of the publication. These supposed ‘misclassifications’ (from the researcher's subjective point of view) can always happen. Therefore, manual evaluation of the clusters and associated decision rules is always required. This manual analysis of clusters additionally enables the researcher to further improve the results by identifying related clusters that can be aggregated or find big clusters that can be split into multiple sub-clusters, or split them if the topics included in the overarching cluster are spread too much vice versa. Since the automatic generation of clusters is based on statistical analysis, the decision criteria might differ from a human interpretation since humans tend to interpret the meaning of topics and they do not solely rely on statistics and logical reasoning when structuring content and when assigning items to that structure. In summary, the artifact provides a supporting mechanism to speed up and standardize the process of literature reviews and increases automation of an otherwise entirely manual process to ultimately improve quality and reproducibility of this important aspect of research.

7.6 Conclusion

In summary, we have developed an artifact based on the word2vec algorithm, LDA topic modeling, rapid automatic keyword extraction and agglomerative hierarchical clustering. This artifact is a first step towards simplifying the task of literature reviews within scientific research.

For this purpose, the publications are first collected by a crawler and vectorized afterwards. Following this, keywords are extracted and the LDA method is used to identify topics. Finally, the word vectors are used to form clusters. These results are presented graphically, for example in the form of dendrograms.

To evaluate the artifact, we used an exemplary set of 308 scientific publications. As the evaluation showed, our extension is particularly suitable for capturing the topic of clusters without looking directly into each paper in detail. However, even in this case there are cluster topics that are not obvious at first glance. Looking at the combination of extracted keywords and topics can help to understand the reason for the clustering. This type of clustering also opens up new perspectives on topics that might be clustered due to other aspects, as might be possible at first appearance.

As every study, also the present study and its results are to be seen and interpreted in consideration of certain limitations. The rather small evaluation data set of just 308 full-text papers which were manually checked if the proposed clustering of our model reflects the expectation of IS researchers, can only serve as a starting point for future research. Another limitation results from the conversion in plain text documents since all information stored in images and figures is not considered by the artifact. Furthermore, a more rigorous evaluation of the artifact's utility for researchers during the creation of literature reviews in different contexts should be subject to future research. In addition, it would be interesting to conduct a comparative performance analysis along different topic modeling approaches such as LDA, latent semantic analysis (LSA), probabilistic LSA (pLSA), etc. and to evaluate possible improvements that can be achieved by optimizing the implemented algorithms.

Finally, this work provides an insight into how the knowledge available in unstructured text data can be efficiently organized and used. This approach might support researchers in conducting comprehensive literature reviews through machine learning.

8 Contributions and Implications

Today, organizations are increasingly using AI—and especially ML—to improve and extend products, services, and processes. Due to the distinctive characteristics of these applications (e.g., data-generated models, uncertainty), many challenges arise for organizations when developing AI-based IS. However, not only organizations are affected, but so are users who are offered services that use AI. Against this background, the overarching objective of this dissertation was to gain a better understanding of how organizations are adopting AI and coping with related challenges, as well as how individual users perceive AI-based IS. Additionally, an ML-based IT artifact was designed to evaluate the capabilities of AI-based IS and their ability to support users. Specifically, the following ROs were the primary focus:

- Analyzing readiness factors for AI adoption in organizations (chapter 3);
- Analyzing success factors for conducting data science competitions (chapter 4);
- Analyzing users' perceptions of the advantages of AI-based advisory systems and the efficacy of trust-increasing mechanisms. (chapter 5);
- Analyzing the utilization of advice with regard to perceived advisor characteristics (chapter 6); and
- Designing an ML-based IT artifact to support structured literature reviews (chapter 7).

8.1 Theoretical Contributions

Overall, the findings of the research papers that represent the core of this dissertation relate to two paradigms that characterize much of IS research: behavioral science and design science (Hevner et al. 2004). The first four research papers belong to the first paradigm and contribute to IS research in the field of IS use by advancing the understanding of organizational adoption of AI (chapters 3 and 4) and individual use of AI-based IS (chapters 5 and 6). The fifth research paper contributes to the second paradigm of design science by creating an IT artifact that uses ML to increase the level of automation of structured literature reviews. Since these studies draw on largely different research and literature streams, the theoretical contributions that can be derived from them are also for the most part discussed separately in the remainder of this section.

Regarding the organizational perspective, Research Paper A (chapter 3) largely confirms existing readiness factors of new technologies but also identifies new factors for the adoption of AI in organizations (e.g., Alsheibani et al. 2018; Crowston and Bolici 2019; Nascimento et al. 2018). Drawing on the TOE framework for technology adoption as a conceptual starting point and using qualitative expert interviews, we show that the general framework is also applicable to the adoption of AI. By extending the framework with newly identified AI-specific factors and subcategories such as data or regulatory restrictions, an adapted framework tailored to the adoption of AI is proposed. This conceptualization can serve as a basis for further research on AI adoption in organizations. For example, subsequent studies could focus on specific factors (e.g., organizational size, organizational culture), specific industries (e.g., healthcare, finance) and associated requirements, or specific business areas (e.g., sales, manufacturing).

The study in Research Paper B (chapter 4) shows that, while hosting data science competitions can be beneficial, organizations should also be aware of the limitations of doing so. For example, only a small portion of the data science process can be represented in these competitions (i.e., model selection and optimization), and the hosting platforms often limit the problem statements to supervised learning problems. Therefore, competitions cannot be used to counter the talent shortage of data scientists experienced by many companies. By identifying factors that affect the success of data science competitions, this explorative study combining data from expert interviews and crawled data can provide a basis for future related research.

The first study in the context of individual use of AI (Research Paper C, chapter 5) applies the concept of relative advantage (Choudhury and Karahanna 2008) from the theory of innovation diffusion (Rogers 2003) to the context of AI-based advisory systems. While the JAS literature considers trust the most influential factor in advice utilization, the study's results show that users value the informational convenience (i.e., availability, instant response) of AI-based advisory systems the most when compared with human advisors. Furthermore, the study confirms the trust-increasing effect of many mechanisms in AI-based IS that have been discussed in previous literature (e.g., Ribeiro et al. 2016; de Visser et al. 2016; Zanker 2012). While previous studies have focused on the effects of specific trust-increasing mechanisms (Hegel et al. 2009; Nilashi et al. 2016), this study compares different mechanisms, and the findings suggest they have varying efficacies for AI-based advisory systems. Specifically, non-binding testing generates the most trust, while adding anthropomorphic features (i.e., visual human representation, speech) had the weakest impact. As we conducted the study in the context of financial robo-advisors, future research could investigate whether the findings also hold true in other contexts and how the perceptions of the relative advantages can be influenced.

While previous studies in the JAS literature stream have investigated the perception of advice from statistical models and “conventional” algorithms (Önkal et al. 2009; Wærn and Ramberg 1996), the results of the second study within the context of individual AI use (Research Paper D, chapter 6) indicate that users might prefer advice from AI-based advisory systems rather than human advisors. A group comparison also showed that different characteristics affect whether users deem a robo- or human advisor suitable for the task. While robo-advisors should demonstrate expertise and efficiency, human advisors should show expertise and integrity. Interestingly, robo-advisors were generally perceived to have higher integrity than their human counterparts. Furthermore, the results show that task–advisor fit is a predictor of users’ advice utilization. Therefore, a connection of the TTF model and the JAS paradigm is feasible. Future research should validate these findings in other tasks related to financial advising as well as in other contexts. In particular, investigating the influence of different task characteristics (e.g., difficulty, locus of control) and evaluating their influence on the task–advisor fit could be highly interesting.

Lastly, the study within the design science context (Research Paper E, chapter 7) proposes an IT artifact to support conducting structured literature reviews using ML. The artifact is based on the word2vec algorithm, LDA topic modeling, rapid automatic keyword extraction, and agglomerative hierarchical clustering. The artifact extends previous research (Dann et al. 2017) by further standardizing and automating the mostly manual process of structured literature reviews with the objective of improving quality and reproducibility. Future research should further validate and evaluate the proposed artifact. Additionally, the artifact itself can be improved—for example, by including content analysis of figures and images, as well as testing different topic modeling or entirely different content analysis approaches.

In sum, this dissertation provides theoretical implications with empirical evidence for the emerging body of research on organizational and individual acceptance and use of AI. The adoption of AI, which is becoming an omnipresent technology, poses a multifaceted and highly complex challenge. By conducting multiple studies, existing research on organizational AI adoption and individual use of AI-based advisory systems was extended. The presented theoretical implications offer a basis for future research to further validate these findings and encourage others to advance research in this field.

8.2 Practical Implications

Aside from the theoretical contributions, various practical implications and recommendations can be derived from the findings of the studies in this dissertation. In particular, managers within organizations, individuals involved in the development of AI-based IS, and the providers of such systems can benefit from the results. Analogous to the procedure in the previous section, the main practical implications are presented separately for each study.

The first study conducted in the organizational context (Research Paper A, chapter 3) showed that a variety of technological, organizational, and environmental factors can enable or impede organizations’ adoption of AI. For example, it is important that business processes are compatible with the characteristics of AI; that data quality, availability, and security are ensured; and that corporate culture and top management are open to innovation. The proposed framework allows companies to evaluate their status quo with regard to AI adoption and helps identify area of improvements to implement AI in their products, services, and processes. Furthermore, general approaches for coping with challenges that often occur when implementing AI were presented. Discrepancies between the expectations of AI providers and customers were revealed, allowing the parties to make arrangements to align their requirements and demands.

The results of the second study within the organizational context (Research Paper B, chapter 4) indicate that crowdsourcing data science through competitions can be beneficial. However, it is important that companies manage their expectations regarding the effort necessary to host such a competition as well as its outcome. Although all technical requirements are taken care of by the platform, the most time-consuming parts of the data science process (i.e., data identification, data aggregation, and data preprocessing) must be performed before launching the competition. Additionally, while in most cases the procedure during data science projects is highly iterative, the competition’s objective and performance measure cannot be altered after launch. With regard to the outcome, organizations should by no means assume that they will receive a holistic solution that can easily be operationalized. However, if the organization’s main objective is to gather novel, innovative approaches to tackle a specific ML problem and to learn from and have professional conversations with the participants, such competitions can be successful.

The results of the first study in the context of individual use of AI (Research Paper C, chapter 5) are beneficial for organizations that offer or are intending to offer AI-based advisory services, such as financial robo-advisors. The results show that users see relative advantages in the use of AI in advisory services compared with traditional human advisors, which reduces uncertainty for providers. Additionally, service providers can leverage the findings by focusing their development and marketing on the relative advantages with the greatest impact. Furthermore, the ranking of trust-increasing mechanisms can help organizations evaluate, select, and prioritize such measures to spend their resources optimally.

The second study within the context of individual use of AI (Research Paper D, chapter 6) provides first evidence that users might choose to utilize AI-based financial advice more than human advice. This indicates that the deployment of robo-advisors can be beneficial for enterprises. Since task–advisor fit appears to predict actual advice utilization, companies could evaluate potential robo-advisory services by conducting market research. Furthermore, providers of such services can leverage the findings regarding the impact of the perceived advisor characteristics by highlighting the robo-advisor’s expertise and its capability to increase the user’s ability to make decisions more efficiently.

The last study (Research Paper E, chapter 7) was conducted in the context of design science and proposes an IT artifact to support scientific structured literature reviews. The artifact addresses a large and coherent part of the structured literature review process, from retrieving research papers from scientific databases to content analysis. The findings suggest that ML methods in the area of natural language processing, such as LDA topic modeling, can meaningfully support content analysis of scientific research papers. Generally, the artifact can be applied to all use cases in which it is helpful that unstructured textual data is clustered by similarity of content (e.g., court rulings on similar lawsuits). Furthermore, the artifact can be seen as a framework that can be altered according to one’s needs (e.g., using alternative preprocessing methods such as bag of words or content analysis techniques like LSA or pLSA.

In conclusion, this dissertation is a step toward understanding the organizational adoption of AI and users’ perceptions and use of AI-based advisory systems. In summary, the studies’ findings provide helpful insights and recommendations for organizations regarding the challenges associated with AI projects, as well as for providers of AI-based advisory systems regarding their customers’ behavior and needs. Ultimately, the findings help organizations succeed in integrating AI in their offerings and thereby gaining an economic and competitive advantage. Nevertheless, further investigations are needed to verify this research and transfer it to different contexts. The research presented in this dissertation is a starting point for further advancing the knowledge on the specifics of the adoption, implementation, and use of AI-based IS.

References

van der Aalst, W. 2016. “Process Mining: Data Science in Action,” *Process Mining: Data Science in Action* (2nd ed.), Berlin Heidelberg: Springer-Verlag.

Ackoff, R. L. 1989. “From Data to Wisdom,” *Journal of Applied Systems Analysis* (16:1), pp. 3–9.

Adadi, A., and Berrada, M. 2018. “Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI),” *IEEE Access* (6), pp. 52138–52160.

Afonso, A. R., and Duque, C. G. 2014. “Automated Text Clustering of Newspaper and Scientific Texts in Brazilian Portuguese: Analysis and Comparison of Methods,” *Journal of Information Systems and Technology Management* (11:2), pp. 415–436.

Afuah, A., and Tucci, C. L. 2012. “Crowdsourcing As a Solution to Distant Search,” *Academy of Management Review* (37:3), pp. 355–375.

Agrawal, A., Gans, J., and Goldfarb, A. 2018. “Is Your Company’s Data Actually Valuable in the AI Era?” (https://hbr.org/2018/01/is-your-companys-data-actually-valuable-in-the-ai-era, accessed November 18, 2018).

Ai, Q., Yang, L., Guo, J., and Croft, W. B. 2016. “Analysis of the Paragraph Vector Model for Information Retrieval,” in *Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval*, Newark, Delaware, USA, pp. 133–142.

Ajzen, I. 1991. “The Theory of Planned Behavior,” *Orgnizational Behavior and Human Decision Processes* (50:2), pp. 179–211.

Akerkar, R. 2013. “Advanced Data Analytics for Business,” *Big Data Computing* (377:9).

Aljaber, B., Stokes, N., Bailey, J., and Pei, J. 2010. “Document Clustering of Scientific Texts Using Citation Contexts,” *Information Retrieval* (13:2), pp. 101–131.

Alsheibani, S., Cheung, Y., and Messom, C. 2018. “Artificial Intelligence Adoption: AI-Readiness at Firm-Level,” in *Proceedings of Pacific Asia Conference on Information Systems (PACIS)*, Yokohama, Japan, A-37.

Andrews, W., Sau, M., Dekate, C., Mullen, A., Brant, K. F., Revang, M., and Plummer, D. C. 2017. “Predicts 2018: Artificial Intelligence”.

Anthes, G. 2017. “Artificial Intelligence Poised to Ride a New Wave,” *Communications of the ACM* (60:7), pp. 19–21.

Arun, R., Suresh, V., Veni Madhavan, C. E., and Narasimha Murthy, M. N. 2010. “On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations,” in *Proceedings of the 14th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, pp. 391–402.

Atsmon, Y., Baroudy, K., Jain, P., Kishore, S., McCarthy, B., Nair, S., and Saleh, T. 2021. “Tipping the Scales in AI: How Leaders Capture Exponential Returns” (https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/tipping-the-scales-in-ai#, accessed May 8, 2021).

Awoyemi, J. O., Adetunmbi, A. O., and Oluwadare, S. A. 2017. “Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis,” in *Proceedings of the IEEE International Conference on Computing, Networking and Informatics*, Ota, Nigeria, pp. 1–9.

Baier, L., Jöhren, F., and Seebacher, S. 2019. “Challenges in the Deployment and Operation of Machine Learning in Practice,” in *Proceedings of the 27th European Conference on Information Systems (ECIS)*, Stockholm & Uppsala, Sweden.

Baskerville, R. L., Kaul, M., and Storey, V. C. 2015. “Genres of Inquiry in Design-Science Research: Justification and Evaluation of Knowledge Production,” *MIS Quarterly* (39:3), pp. 541–564.

Bayardo, R. J., and Agrawal, R. 2005. “Data Privacy Through Optimal K-Anonymization,” in *Proceedings of the 21st International Conference on Data Engineering*.

Beketov, M., Lehmann, K., and Wittke, M. 2018. “Robo Advisors: Quantitative Methods inside the Robots,” *Journal of Asset Management* (19), pp. 363–370.

Berente, N., Gu, B., Recker, J., and Santhanam, R. 2019. “Call for Papers MISQ Special Issue on Managing AI,” *MIS Quarterly*.

Berkhin, P. 2006. “A Survey of Clustering Data Mining Techniques,” in *Grouping Multidimensional Data*, Berlin Heidelberg: Springer-Verlag, pp. 25–71.

Berner, C., Brockman, G., Chan, B., Cheung, V., Dębiak, P., Dennison, C., Farhi, D., Fischer, Q., Hashme, S., Hesse, C., Józefowicz, R., Gray, S., Olsson, C., Pachocki, J., Petrov, M., de Oliveira Pinto, H. P., Raiman, J., Salimans, T., Schlatter, J., Schneider, J., Sidor, S., Sutskever, I., Tang, J., Wolski, F., and Zhang, S. 2019. “Dota 2 with Large Scale Deep Reinforcement Learning,” *OpenAI* (https://arxiv.org/abs/1912.06680).

Bhattacherjee, A., and Premkumar, G. 2004. “Understanding Changes in Belief and Attitude Toward Information Technology Usage: A Theoretical Model and Longitudinal Test,” *MIS Quarterly* (28:2), pp. 229–254.

Bichler, M., Heinzl, A., and van der Aalst, W. 2017. "Business Analytics and Data Science: Once Again?," *Business & Information Systems Engineering* (59:2), pp. 77–79.

Bishop, C. M. 2006. *Pattern Recognition and Machine Learning*, New York, NY, USA: Springer Science and Business Media.

Blei, D., Carin, L., and Dunson, D. 2010. "Probabilistic Topic Models: A Focus on Graphical Model Design and Applications to Document and Image Analysis," *IEEE Signal Processing Magazine* (27:6), pp. 55–65.

Blei, D. M. 2012. "Probabilistic Topic Models," *Communications of the ACM* (55:4), p. 77.

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

Bonaccio, S., and Dalal, R. S. 2006. "Advice Taking and Decision-Making: An Integrative Literature Review, and Implications for the Organizational Sciences," *Organizational Behavior and Human Decision Processes* (101:2), pp. 127–151.

Bornmann, L., and Mutz, R. 2015. "Growth Rates of Modern Science: A Bibliometric Analysis Based on the Number of Publications and Cited References," *Journal of the Association for Information Science and Technology* (66:11), pp. 2215–2222.

Bremser, C. 2018. "Starting Points for Big Data Adoption," in *Proceedings of European Conference on Information Systems (ECIS)*, Portsmouth, United Kingdom, A-32.

Bremser, C., Piller, G., and Rothlauf, F. 2017. "Strategies and Influencing Factors for Big Data Exploration," in *Proceedings of Americas Conference on Information Systems (AMCIS)*, Massachusetts, USA, August 10, A-4.

vom Brocke, J., Maaß, W. M., Buxmann, P., Maedche, A., Leimeister, J. M., and Pecht, G. 2018. "Future Work and Enterprise Systems," *Business & Information Systems Engineering* (60:4), pp. 357–366.

Brocke, J. vom, Simons, A., Niehaves, B., Reimer, K., Plattfaut, R., and Cleven, A. 2009. "Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process," in *ECIS 2009 Proceedings*, Paper 161.

Brook, J. S., Whiteman, M., and Cohen, P. 1995. "Stage of Drug Use, Aggression, and Theft/Vandalism: Shared and Unshared Risks," in *Drugs, Crime, and Other Deviant Adaptions: Longitudinal Studies*, H. B. Kaplan (ed.), New York: Springer Science and Business Media, pp. 83–96.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., HÉigeartaigh, S. Ó., Beard, S., Belfield, H., Farquhar, S., Lyle, C., Crootof, R., Evans, O., Page, M., Bryson, J., Yampolskiy, R., and Amodei, D. 2018. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*.

Brynjolfsson, E., Hitt, L. M., and Kim, H. H. 2011. "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?," *SSRN Electronic Journal*.

Brynjolfsson, E., and McAfee, A. 2017. "The Business of Artificial Intelligence," *Harvard Business Review* (https://hbr.org/2017/07/the-business-of-artificial-intelligence, accessed April 26, 2021).

Brynjolfsson, E., and Mitchell, T. 2017. "What Can Machine Learning Do? Workforce Implications," *Science* (358:6370), pp. 1530–1534.

Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., Henke, N., and Trench, M. 2017. *Artificial Intelligence: The Next Digital Frontier?*, Technical Report. McKinsey Global Institute.

Burkhard, R. A. 2004. "Learning from Architects: The Difference between Knowledge Visualization and Information Visualization," in *Proceedings of the International Conference on Information Visualization* (Vol. 8), pp. 519–524.

Burrell, J. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms," *Big Data & Society* (3:1), pp. 1–12.

Burton-Jones, A., and Gallivan, M. J. 2007. "Towards a Deeper Understanding of System Usage in Organizations: A Multilevel Perspective," *MIS Quarterly* (31:4), pp. 657–679.

Burton-Jones, A., Stein, M.-K., and Mishra, A. 2017. "IS Use," in *MIS Quarterly Research Curations*, A. Bush and A. Rai (eds.).

Camerer, C. F., and Hogarth, R. M. 1999. "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework," *Journal of Risk and Uncertainty* (19:1/3), pp. 7–42.

Cao, J., Xia, T., Li, J., Zhang, Y., and Tang, S. 2009. "A Density-Based Method for Adaptive LDA Model Selection," *Neurocomputing* (72:7–9), pp. 1775–1781.

Castelo, N., Bos, M. W., and Lehmann, D. R. 2019. "Task-Dependent Algorithm Aversion," *Journal of Marketing Research* (56:5), pp. 809–825.

Chan, Y. E., Huff, S. L., Barclay, D. W., and Copeland, D. G. 1997. "Business Strategic Orientation, Information Systems Strategic Orientation, and Strategic Alignment," *Information Systems Research* (8:2), pp. 125–150.

Chandler, D., and Kapelner, A. 2013. "Breaking Monotony with Meaning: Motivation in Crowdsourcing Markets," *Journal of Economic Behavior & Organization* (90), pp. 123–133.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. 2000. *CRISP-DM 1.0 Step-by-Step Data Mining Guide*.

Chen, H., Chiang, R. H. L., and Storey, V. C. 2012. "Business Intelligence and Analytics: From Big Data to Big Impact," *MIS Quarterly* (36:4), pp. 1165–1188.

Chesbrough, H., Vanhaverbeke, W., and West, J. (eds.) 2006. *Open Innovation: Researching a New Paradigm*, Oxford University Press on Demand.

Choudhury, V., and Karahanna, E. 2008. "The Relative Advantage of Electronic Channels: A Multidimensional View," *MIS Quarterly* (32:1), pp. 179–200.

Corbin, J., and Strauss, A. 2014. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, Thousand Oaks: Sage publications.

Córdoba, J.-R., Pilkington, A., and Bernroider, E. W. N. 2012. "Information Systems as a Discipline in the Making: Comparing EJIS and MISQ between 1995 and 2008," *European Journal of Information Systems* (21:5), pp. 479–495.

Costello, K. 2019. "Gartner Survey Shows 37 Percent of Organizations Have Implemented AI in Some Form" (https://www.gartner.com/en/newsroom/press-releases/2019-01-21-gartner-survey-shows-37-percent-of-organizations-have, accessed May 14, 2019).

Crowston, K., and Bolici, F. 2019. "Impacts of Machine Learning on Work," in *Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS),* Wailea, Hawaii, USA, pp. 5961–5970.

Daft, R. L., Lengel, R. H., and Trevino, L. K. 1987. "Message Equivocality, Media Selection, and Manager Performance: Implications for Information Systems," *MIS Quarterly* (11:3), pp. 355–366.

Dann, D., Hauser, M., and Hanke, J. 2017. "Reconstructing the Giant: Automating the Categorization of Scientific Articles with Deep Learning Techniques," in *Proceedings Der 13. Internationalen Tagung Wirtschaftsinformatik*, St. Gallen, pp. 1538–1549.

Darrat, A. A., Darrat, M. A., and Amyx, D. 2016. "How Impulse Buying Influences Compulsive Buying: The Central Role of Consumer Anxiety and Escapism," *Journal of Retailing and Consumer Services* (31), pp. 103–108.

Davenport, T. H., and Patil, D. J. 2012. "Data Scientist: The Sexiest Job of the 21st Century," *Harvard Business Review* (90:5), pp. 70–76.

Davis, F. D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* (13:3), pp. 319–340.

Davison, A. C., and Hinkley, D. V. 1997. *Bootstrap Methods and Their Application*, (10th ed.), New York, NY: Cambridge University Press.

Deci, E. L., and Ryan, R. M. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*, Boston, MA: Springer US.

DePietro, R., Wiarda, E., and Fleischer, M. 1990. “The Context for Change: Organization, Technology and Environment,” in *The Process of Technological Innovation* (4th ed.), L. Tornatzky and M. Fleischer (eds.), Lexington: Lexington Books, pp. 152–175.

Devaraj, S., Fan, M., and Kohli, R. 2002. “Antecedents of B2C Channel Satisfaction and Preference: Validating e-Commerce Metrics,” *Information Systems Research* (13:3), pp. 316–333.

Deveaud, R., SanJuan, E., and Bellot, P. 2014. “Accurate and Effective Latent Concept Modeling for Ad Hoc Information Retrieval,” *Document Numérique* (17:1), pp. 61–84.

Dhar, V. 2013. “Data Science and Prediction,” *Communications of the ACM* (56:12), pp. 64–73.

Dietvorst, B. J., Simmons, J. P., and Massey, C. 2015. “Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err,” *Journal of Experimental Psychology: General* (144:1), pp. 114–126.

Dumouchel, B., and Demaine, J. 2006. “Knowledge Discovery in the Digital Library: Access Tools for Mining Science,” *Information Services & Use* (26), pp. 29–44.

Dzielinski, M. 2012. “Measuring Economic Uncertainty and Its Impact on the Stock Market,” *Finance Research Letters* (9:3), pp. 167–175.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., and Beck, H. P. 2003. “The Role of Trust in Automation Reliance,” *International Journal of Human Computer Studies* (58:6), pp. 697–718.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., and Kruger, J. 2008. “Why the Unskilled Are Unaware: Further Explorations of (Absent) Self-Insight Among the Incompetent.,” *Organizational Behavior and Human Decision Processes* (105:1), pp. 98–121.

Eickhoff, C. 2014. “Crowd-Powered Experts: Helping Surgeons Interpret Breast Cancer Images,” in *Proceedings of the First International Workshop on Gamification for Information Retrieval*, Amsterdam, The Netherlands, pp. 53–56.

Eickhoff, M., Muntermann, J., and Weinrich, T. 2017. “What Do FinTechs Actually Do? A Taxonomy of FinTech Business Models,” in *Proceedings of the 38th International Conference on Information Systems*, Seoul, pp. 1–19.

Elliot, B., and Andrews, W. 2017. "A Framework for Applying AI in the Enterprise," *Gartner*, pp. 1–38 (https://www.gartner.com/ngw/globalassets/en/information-technology/documents/insights/a-framework-for-applying-ai-in-the-enterprise.pdf, accessed November 26, 2018).

Ericsson 2020. "Adopting AI in Organizations" (https://www.ericsson.com/4ab2b3/assets/local/reports-papers/industrylab/doc/adopting-ai-report.pdf, accessed April 24, 2021).

Estellés-Arolas, E., and González-Ladrón-de-Guevara, F. 2012. "Towards an Integrated Crowdsourcing Definition," *Journal of Information Science* (38:2), pp. 189–200.

Eule, A. 2017. "Rating the Robo-Advisors," *Barron's - The Dow Jones Business and Financial Weekly* (https://www.barrons.com/articles/rating-the-robo-advisors-1501303316, accessed March 1, 2019).

Eurostat 2018. "Internet Access and Use Statistics - Households and Individuals".

Ezell, S., and Marxgut, P. 2015. "Comparing American and European Innovation Cultures," *ITIF*, pp. 157–199 (http://www2.itif.org/2015-comparing-american-european-innovation-cultures.pdf, accessed March 16, 2019).

Feng, B., and MacGeorge, E. L. 2006. "Predicting Receptiveness to Advice: Characteristics of the Problem, the Advice-Giver, and the Recipient," *Southern Communication Journal* (71:1), pp. 67–85.

Fishbein, M., and Ajzen, I. 1975. *Belief, Attitude, Intention and Behaviour: An Introduction to Theory and Research*, Reading, MA: Addison-Wesley.

Flam, F. 2018. "IBM's Watson Hasn't Beaten Cancer, But A.I. Still Has Promise" (https://www.bloomberg.com/opinion/articles/2018-08-24/ibm-s-watson-failed-against-cancer-but-a-i-still-has-promise, accessed November 20, 2018).

Fleming, N. 2018. "How Artificial Intelligence Is Changing Drug Discovery," *Nature* (557:7706), S55+.

Flick, U. 2004. "No Triangulation in Qualitative Research," in *A Companion to Qualitative Research*, U. Flick, E. von Kardorff, and I. Steinke (eds.), London: Sage, pp. 178–183.

Flick, U., Kardorff, E., and Steinke, I. 2004. "What Is Qualitative Research? An Introduction to the Field," in *A Companion to Qualitative Research*, U. Flick, E. Kardoff, and I. Steinke (eds.), London: Sage, pp. 3–11.

Fornell, C., and Bookstein, F. L. 1982. "Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory," *Journal of Marketing Research* (19:4), pp. 440–452.

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

Frambach, R. T., and Schillewaert, N. 2002. "Organizational Innovation Adoption: A Multi-Level Framework of Determinants and Opportunities for Future Research," *Journal of Business Research* (55:2), pp. 163– 176.

Frey, F. J., and Luks, M. 2016. "The Innovation-Driven Hackathon," in *Proceedings of the 21st European Conference on Pattern Languages of Programs*, New York, NY, USA, pp. 1–11.

Garcia Martinez, M. 2015. "Solver Engagement in Knowledge Sharing in Crowdsourcing Communities: Exploring the Link to Creativity," *Research Policy* (44:8), pp. 1419–1430.

Gartner 2020. "Gartner Identifies the Top Strategic Technology Trends for 2021" (https://www.gartner.com/en/newsroom/press-releases/2020-10-19-gartner-identifies-the-top-strategic-technology-trends-for-2021, accessed May 9, 2021).

Gedikli, F., Jannach, D., and Ge, M. 2014. "How Should i Explain? A Comparison of Different Explanation Types for Recommender Systems," *International Journal of Human Computer Studies* (72:4), pp. 367–382.

George, S., and Joseph, S. 2014. "Text Classification by Augmenting Bag of Words (BOW) Representation with Co-Occurrence Feature," *IOSR Journal of Computer Engineering* (16:1), pp. 34–38.

Gino, F., and Moore, D. A. 2007. "Effects of Task Difficulty on Use of Advice," *Journal of Behavioral Decision Making* (20:1), pp. 21–35.

Gino, F., Shang, J., and Croson, R. 2009. "The Impact of Information from Similar or Different Advisors on Judgment," *Organizational Behavior and Human Decision Processes* (108:2), pp. 287–302.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. 2014. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587.

Gnewuch, U., Morana, S., and Mädche, A. 2017. "Towards Designing Cooperative and Social Conversational Agents for Customer Service," in *Proceedings of the International Conference on Information Systems (ICIS) 2017*, Seoul, South Korea, December 10.

Goasduff, L. 2020. "How to Staff Your AI Team" (https://www.gartner.com/smarterwithgartner/how-to-staff-your-ai-team/, accessed May 10, 2021).

Goeke, M. 2016. "Kompetenz Und Trends Im Private Banking," in *Banking & Innovation 2016*, Wiesbaden: Springer Fachmedien Wiesbaden, pp. 3–9.

Gönül, M. S., Önkal, D., and Lawrence, M. 2006. "The Effects of Structural Characteristics of Explanations on Use of a DSS," *Decision Support Systems* (42:3), pp. 1481–1493.

Goodfellow, I., Bengio, Y., and Courville, A. 2016. *Deep Learning*, Cambridge, MA, USA: The MIT press.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. 2014. "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems 27*.

Goodhue, D. L. 1995. "Understanding User Evaluations of Information Systems," *Management Science* (41:12), pp. 1827–1844.

Goodhue, D. L., Klein, B. D., and March, S. T. 2000. "User Evaluations of IS as Surrogates for Objective Performance," *Information & Management* (38:2), pp. 87–101.

Goodhue, D. L., and Thompson, R. L. 1995a. "Task-Technology Fit and Individual Performance," *MIS Quarterly* (19:2), pp. 213–236.

Goodhue, D. L., and Thompson, R. L. 1995b. "Task-Technology Fit and Individual Performance," *MIS Quarterly* (19:2), p. 213.

Greener, S. 2008. *Business Research Methods*, London: Ventus Publishing ApS.

Greenhouse, S. W., and Geisser, S. 1959. "On Methods in the Analysis of Profile Data," *Psychometrika* (24:2), pp. 95–112.

Gregor, K., Danihelka, I., Graves, A., Rezende, D., and Wierstra, D. 2015. "Draw: A Recurrent Neural Network for Image Generation," in *International Conference on Machine Learning*, pp. 1462–1471.

Gregor, S. 2001. "Explanations from Knowledge-Based Systems and Cooperative Problem Solving: An Empirical Study," *International Journal of Human Computer Studies* (54:1), pp. 81–105.

Griffiths, T. L., and Steyvers, M. 2004. "Finding Scientific Topics," *Proceedings of the National Academy of Sciences of the United States of America* (101:suppl 1), pp. 5228–35.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. 2018. "A Survey of Methods for Explaining Black Box Models," *ACM Computing Surveys (CSUR)* (51:5), pp. 1–42.

Gulo, C. A. S. J., Rúbio, T. R. P. M., Tabassum, S., and Prado, S. G. D. 2015. "Mining Scientific Articles Powered by Machine Learning Techniques," *2015 Imperial College Computing Student Workshop (ICCSW 2015)* (49), pp. 21–28.

Günther, E., and Quandt, T. 2015. "Word Counts and Topic Models," *Digital Journalism* (4:1), pp. 75–88.

Hair, J. F., Ringle, C. M., and Sarstedt, M. 2011. "PLS-SEM: Indeed a Silver Bullet," *Journal of Marketing Theory and Practice* (19:2), pp. 139–152.

Hair, J. J. F., Hult, G. T. M., Ringle, C., and Sarstedt, M. 2013. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, London: Sage Publications.

Harvey, N., and Fischer, I. 1997. "Taking Advice: Accepting Help, Improving Judgment, and Sharing Responsibility," *Organizational Behavior and Human Decision Processes* (70:2), pp. 117–133.

Hatta, N. N. M., Miskon, S., and Abdullah, N. S. 2017. "Business Intelligence System Adoption Model for SMEs," in *Proceedings of Pacific Asia Conference on Information Systems (PACIS)*, Langkawi, Malaysia, A-192.

Hegel, F., Lohse, M., and Wrede, B. 2009. "Effects of Visual Appearance on the Attribution of Applications in Social Robotics," in *The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 64–71.

Hein, D., Rauschnabel, P., He, J., Richter, L., and Ivens, B. 2018. "What Drives the Adoption of Autonomous Cars?," in *ICIS 2018 Proceedings*, San Francisco, USA, December 13.

Hermanns, H. 2004. "Interviewing as an Activity," in *A Companion to Qualitative Research*, U. Flick, E. von Kardoff, and I. Steinke (eds.), London: Sage, pp. 209–213.

Hertel, G., Niedner, S., and Herrmann, S. 2003. "Motivation of Software Developers in Open Source Projects: An Internet-Based Survey of Contributors to the Linux Kernel," *Research Policy* (32:7), pp. 1159–1177.

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

Hill, J., Ford, W. R., and Farreras, I. G. 2015. "Real Conversations with Artificial Intelligence: A Comparison between Human-Human Online Conversations and Human-Chatbot Conversations," *Computers in Human Behavior* (49), pp. 245–250.

von Hippel, E. 2005. *Democratizing Innovation: Users Take Center Stage*, Cambridge, MA: MIT Press.

Hosanagar, K., and Saxena, A. 2017. "The First Wave of Corporate AI Is Doomed to Fail" (https://hbr.org/2017/04/the-first-wave-of-corporate-ai-is-doomed-to-fail, accessed November 18, 2018).

Howe, J. 2006. "The Rise of Crowdsourcing," *Wired Magazine* (14:6), pp. 1–4.

HSBC 2017. "Trust in Technology," pp. 1–32 (https://www.hsbc.com/-/files/hsbc/media/media-release/2017/170609-updated-trust-in-technology-final-report.pdf).

Hsieh, H. F., and Shannon, S. E. 2005. “Three Approaches to Qualitative Content Analysis,” *Qualitative Health Research* (15:9), pp. 1277–1288.

Hu, L., and Bentler, P. M. 1999. “Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives,” *Structural Equation Modeling: A Multidisciplinary Journal* (6:1), pp. 1–55.

Huang, M. H., and Rust, R. T. 2018. “Artificial Intelligence in Service,” *Journal of Service Research* (21:2), pp. 155–172.

Huang, W., Nakamori, Y., Wang, S., and Ma, T. 2005. “Mining Scientific Literature to Predict New Relationships,” *Intelligent Data Analysis* (9), pp. 219–234.

Jeppesen, L. B., and Lakhani, K. R. 2010. “Marginality and Problem-Solving Effectiveness in Broadcast Search,” *Organization Science* (21:5), pp. 1016–1033.

Johnson, M., Schuster, M., Le, Q. V, Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F., Wattenberg, M., and Corrado, G. 2017. “Google’s Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation,” *Transactions of the Association for Computational Linguistics* (5), pp. 339–351.

Jordan, M. I., and Mitchell, T. M. 2015. “Machine Learning: Trends, Perspectives, and Prospects,” *Science* (349:6245), pp. 255–260.

Jung, D., Dorner, V., Glaser, F., and Morana, S. 2018. “Robo-Advisory: Digitalization and Automation of Financial Advisory,” *Business and Information Systems Engineering* (60:1), pp. 81–86.

Jung, D., Dorner, V., Weinhardt, C., and Pusmaz, H. 2018. “Designing a Robo-Advisor for Risk-Averse, Low-Budget Consumers,” *Electronic Markets* (28:3), pp. 367–380.

Jung, D., and Weinhardt, C. 2018. “Robo-Advisors and Financial Decision Inertia : How Choice Architecture Helps to Reduce Inertia in Financial Planning Tools,” in *Proceesdings of the 39th International Conference on Information Systems*, San Francisco, USA, pp. 1–17.

Jungermann, H. 1999. “Advice Giving and Taking,” in *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences*, p. 11.

Jussupow, E., and Benbasat, I. 2020. “Why Are We Averse Towards Algorithms? A Comprehensive Literature Review on Algorithm Aversion,” in *Proceedings of the 28th European Conference on Information Systems (ECIS)*, An Online AIS Conference.

Kanawaday, A., and Sane, A. 2018. “Machine Learning for Predictive Maintenance of Industrial Machines Using IoT Sensor Data,” in *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences*, pp. 87–90.

Kaufmann, N., Schulze, T., and Veit, D. 2011. “More than Fun and Money. Worker Motivation in Crowdsourcing-A Study on Mechanical Turk,” in *Proceedings of the Seventeenth Americas Conference on Information Systems*, Detroit, USA.

Kim, H., Benbasat, I., and Cavusoglu, H. 2017. “Online Consumers’ Attribution of Inconsistency Between Advice Sources,” in *Thirty Eighth International Conference on Information Systems*, Seoul, South Korea, pp. 1–10.

Kleemann, F., Voß, G. G., and Rieder, K. 2008. “Un (Der) Paid Innovators: The Commercial Utilization of Consumer Work through Crowdsourcing,” *Science, Technology & Innovation Studies* (4:1), pp. 5–26.

Klopping, I. M., and Mckinney, E. 2004. “Extending the Technology Acceptance Model Extending the Technology Acceptance Model and the Task and the Task-Technology Fit Model to Technology Fit Model to Consumer E Consumer E-Commerce Commerce,” *Information Technology, Learning, and Performance Journal* (22:1), pp. 35–48.

Kober, J., Bagnell, J. A., and Peters, J. 2013. “Reinforcement Learning in Robotics: A Survey,” *The International Journal of Robotics Research* (32:11), pp. 1238–1274.

Komiak, S. Y. X., and Benbasat, I. 2006. “The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents,” *MIS Quarterly* (30:4), pp. 941–960.

Krasnova, H., Spiekermann, S., Koroleva, K., and Hildebrand, T. 2010. “Online Social Networks: Why We Disclose,” *Journal of Information Technology* (25), pp. 109–125.

Kriegel, H.-P., Schubert, E., and Zimek, A. 2017. “The (Black) Art of Runtime Evaluation: Are We Comparing Algorithms or Implementations?,” *Knowledge and Information Systems* (52:2), pp. 341–378.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. 2012. “Imagenet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems 25*, pp. 1097–1105.

Kruger, J., and Dunning, D. 1999. “Unskilled and Unaware of It: How Difficulties in Recognizing One’s Own Incompetence Lead to Inflated Self-Assessments.,” *Journal of Personality and Social Psychology* (77:6), pp. 1121–34.

Krumeich, J., Werth, D., and Loos, P. 2016. “Prescriptive Control of Business Processes: New Potentials Through Predictive Analytics of Big Data in the Process Manufacturing Industry,” *Business and Information Systems Engineering* (58:4), pp. 261–280.

Kruse, L., Wunderlich, N., and Beck, R. 2019. “Artificial Intelligence for the Financial Service Industry: What Challenges Organizations to Succeed,” in *Proceedings of Hawaii International Conference on System Sciences*, Wailea, Hawaii, USA, p. 6408.

Kwon, B. C., Choi, M. J., Kim, J. T., Choi, E., Kim, Y. Bin, Kwon, S., Sun, J., and Choo, J. 2019. "RetainVis: Visual Analytics with Interpretable and Interactive Recurrent Neural Networks on Electronic Medical Records," *IEEE Transactions on Visualization and Computer Graphics* (25:1), pp. 299–309.

De Laat, P. B. 2018. "Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?," *Philosophy & Technology* (31), pp. 525–541.

Lacity, M. C., and Janson, M. A. 1994. "Understanding Qualitative Data: A Framework of Text Analysis Methods," *Journal of Management Information Systems* (11:2), pp. 137–155.

Le, Q., and Mikolov, T. 2014. "Distributed Representations of Sentences and Documents," *Proceedings of the 31st International Conference on Machine Learning*, (P. X. Eric and J. Tony, eds.), pp. 1188–1196.

LeCun, Y., Bengio, Y., and Hinton, G. 2015. "Deep Learning," *Nature* (521:7553), pp. 436–444.

Lee, L. D., and See, K. A. 2004. "Trust in Automation: Designing for Appropriate Reliance," *Human Factors* (24:1), pp. 50–80.

Lee, M.-C. 2009. "Factors Influencing the Adoption of Internet Banking: An Integration of TAM and TPB with Perceived Risk and Perceived Benefit," *Electronic Commerce Research and Applications* (8:3), pp. 130–141.

Leimeister, J. M., Huber, M., Bretschneider, U., and Krcmar, H. 2009. "Leveraging Crowdsourcing: Activation-Supporting Components for IT-Based Ideas Competition," *Journal of Management Information Systems* (26:1), pp. 197–224.

Leonard-Barton, D. 1992. "Core Capabilities and Core Rigidities: A Paradox in Managing New Product Development," *Strategic Management Journal* (13:S1), pp. 111–125.

Lepenioti, K., Bousdekis, A., Apostolou, D., and Mentzas, G. 2020. "Prescriptive Analytics: Literature Review and Research Challenges," *International Journal of Information Management*, pp. 57–70.

Letham, B., Rudin, C., McCormick, T. H., and Madigan, D. 2015. "Interpretable Classifiers Using Rules and Bayesian Analysis: Building a Better Stroke Prediction Model," *Annals of Applied Statistics* (9:3), pp. 1350–1371.

Li, H., Kuo, C., and Rusell, M. G. 2006. "The Impact of Perceived Channel Utilities, Shopping Orientations, and Demographics on the Consumer's Online Buying Behavior," *Journal of Computer-Mediated Communication* (5:2).

Li, N., Li, T., and Venkatasubramanian, S. 2007. "T-Closeness: Privacy Beyond k-Anonymity and l-Diversity," in *23rd International Conference on Data Engineering*, pp. 106–115.

Lian, J.-W., Yen, D. C., and Wang, Y.-T. 2014. "An Exploratory Study to Understand the Critical Factors Affecting the Decision to Adopt Cloud Computing in Taiwan Hospital," *International Journal of Information Management* (34:1), pp. 28–36.

Lin, H.-F. 2011. "An Empirical Investigation of Mobile Banking Adoption: The Effect of Innovation Attributes and Knowledge-Based Trust," *International Journal of Information Management* (31:3), pp. 252–260.

Lipton, Z. C. 2016. "The Mythos of Model Interpretability," in *ICML Workshop on Human Interpretability in Machine Learning*, New York.

Lowry, P. B., D'Arcy, J., Hammer, B., and Moody, G. D. 2016. "'Cargo Cult' Science in Traditional Organization and Information Systems Survey Research: A Case for Using Nontraditional Methods of Data Collection, Including Mechanical Turk and Online Panels," *Journal of Strategic Information Systems* (25:3), pp. 232–240.

Lowry, P. B., Gaskin, J., Twyman, N., Hammer, B., and Roberts, T. 2012. "Taking 'Fun and Games' Seriously: Proposing the Hedonic-Motivation System Adoption Model (HMSAM)," *Journal of the Association for Information Systems* (14:11), pp. 617–671.

Lu, B. Y., Qian, D., Fu, H., and Chen, W. 2018. "Will Supercomputers Be Super-Data and Super-AI Machines?," *Communications of the ACM* (61:11), pp. 82–87.

Lundberg, S. M., and Lee, S. I. 2017. "A Unified Approach to Interpreting Model Predictions," in *Proceedings of Advances in Neural Information Processing Systems*, Long Beach, CA, United States, pp. 4768–4777.

Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M. 2006. "L-Diversity: Privacy beyond k-Anonymity," in *22nd International Conference on Data Engineering (ICDE'06)*, pp. 24–24.

Madhulatha, T. S. 2012. "An Overview on Clustering Methods," *IOSR Journal of Engineering* (2:4), pp. 719–725.

Mallmann, G. L., and Gastaud Maçada, A. C. 2018. "Adoption of Cloud Computing: A Study with Public and Private Hospitals in a Developing Country," *International Journal of Innovation and Technology Management* (15:05), pp. 1–20.

Martens, D., and Provost, F. 2014. "Explaining Data-Driven Document Classification," *MIS Quarterly* (38:1), pp. 73–99.

Mayer, R. C., Davis, J. H., and Schoorman, F. D. 1995. "An Integrative Model of Organizational Trust," *The Academy of Management Review* (20:3), pp. 709–734.

Mayring, P. 2004. "Qualitative Content Analysis," in *A Companion to Qualitative Research* (1st ed.), U. Flick, E. Kardorff, and I. Steinke (eds.), London: Sage Publications, pp. 266–269.

McAfee, A., and Brynjolfsson, E. 2012. "Big Data: The Management Revolution," *Harvard Business Review* (90:10), pp. 60–68.

McCarthy, J. 2007. "WHAT IS ARTIFICIAL INTELLIGENCE?," *Stanford University*, pp. 1–15 (http://jmc.stanford.edu/articles/whatisai/whatisai.pdf, accessed November 26, 2018).

McKnight, D. H., Choudhury, V., and Kacmar, C. 2002. "Developing and Validating Trust Measure for E-Commerce: An Integrative Typology," *Information Systems Research* (13:3), pp. 334–359.

Meade, A. W., and Craig, S. B. 2012. "Identifying Careless Responses in Survey Data," *Psychological Methods* (17:3), pp. 437–455.

Mesbah, N., Tauchert, C., and Buxmann, P. 2021. "Whose Advice Counts More – Man or Machine? An Experimental Investigation of AI-Based Advice Utilization," in *Proceedings of the 54th Hawaii International Conference on System Sciences*, Online Conference.

Mesbah, N., Tauchert, C., Olt, C. M., and Buxmann, P. 2019. "Promoting Trust in AI-Based Expert Systems," in *Proceedings of the 25th Americas Conference on Information Systems*, Cancun, Mexico, August 14.

Metz, R. 2013. "A Startup Called Kaggle Tries to Bring Smart People to Knotty Problems.," *MIT Technology Review* (116:5), p. 51.

Mihalcea, R., and Tarau, P. 2004. "TextRank: Bringing Order into Texts," in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, Spain, pp. 404–411.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. 2013. "Efficient Estimation of Word Representations in Vector Space," pp. 1–12 (https://arxiv.org/abs/1301.3781v3, accessed May 14, 2019).

Miles, M. B., and Huberman, A. M. 1994. *Qualitative Data Analysis: An Expanded Sourcebook*, (2nd ed.), Beverly Hills, CA: Sage.

Miller, D. D., and Brown, E. W. 2018. "Artificial Intelligence in Medical Practice: The Question to the Answer?," *American Journal of Medicine* (131:2), pp. 129–133.

Mitchell, T. M. 1997. *Machine Learning*, New York, NY, USA: McGraw-Hill.

Moncrief, W. C. 2017. "Are Sales as We Know It Dying ... or Merely Transforming?," *Journal of Personal Selling & Sales Management* (37:4), pp. 271–279.

Monroe, D. 2018. "Chips for Artificial Intelligence," *Communications of the ACM* (61:4), pp. 15–17.

Moore, G. C., and Benbasat, I. 1991. "Development of Instrument to Measure the Perceptions of Adopting an Information Technology Innovation," *Information Systems Research* (2:3), pp. 192–222.

Mou, Y., and Xu, K. 2017. "The Media Inequality: Comparing the Initial Human-Human and Human-AI Social Interactions," *Computers in Human Behavior* (72), pp. 432–440.

MSV, J. 2018. "Here Are Three Factors That Accelerate The Rise Of Artificial Intelligence," *Forbes* (https://www.forbes.com/sites/janakirammsv/2018/05/27/here-are-three-factors-that-accelerate-the-rise-of-artificial-intelligence/#208c4888add9, accessed November 23, 2018).

Myers, M. D. 1997. "Qualitative Research in Information Systems.," *MIS Quarterly* (21:2).

Nair, R., and Narayanan, A. 2012. "Data and Technology Perspective Benefitting from Big Data Leveraging Unstructured Data Capabilities for Competitive Advantage," *Booz&Co* (http://assets.fiercemarkets.net/public/sites/energy/reports/BoozCo_Benefitting-from-Big-Data.pdf, accessed June 15, 2019).

Nascimento, A. M., Cunha, M. A. V. C., Souza Meirelles, F., Scornavacca, E., and Melo, V. V 2018. "A Literature Analysis of Research on Artificial Intelligence in Management Information System (MIS)," in *Proceedings of Americas Conference on Information Systems (AMCIS)*, New Orleans, USA, A-5.

Nikita, M. 2019. "Select Number of Topics for LDA Model" (https://cran.r-project.org/web/packages/ldatuning/vignettes/topics.html, accessed May 10, 2019).

Nilashi, M., Jannach, D., Ibrahim, O. bin, Esfahani, M. D., and Ahmadi, H. 2016. "Recommendation Quality, Transparency, and Website Quality for Trust-Building in Recommendation Agents," *Electronic Commerce Research and Applications* (19), pp. 70–84.

Önkal, D., Goodwin, P., Thomson, M., Gönül, S., and Pollock, A. 2009. "The Relative Influence of Advice from Human Experts and Statistical Methods," *Journal of Behavioral Decision Making* (22:4), pp. 390–409.

Palmer, M. 2006. "Data Is the New Oil," *Association of National Advertisers* (https://ana.blogs.com/maestros/2006/11/data_is_the_new.html, accessed April 15, 2019).

Panetta, K. 2017. "Top Trends in the Gartner Hype Cycle for Emerging Technologies, 2017" (https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/, accessed April 24, 2021).

Parkes, A. 2013. "The Effect of Task–Individual–Technology Fit on User Attitude and Performance: An Experimental Investigation," *Decision Support Systems* (54:2), pp. 997–1009.

Pavlou, P. A. 2018. "Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model," *International Journal of Electronic Commerce* (7:3), pp. 101–134.

Petter, S., William, H. D., and McLean, E. R. 2013. "Information System Success: The Quest for Independent Variables," *Journal of MIS* (29:4), pp. 7–62.

Phoon, K., and Koh, F. 2018. "Robo-Advisors and Wealth Management," *Journal of Alternative Investments* (20:3), pp. 79–94.

Pickard, M. D., Roster, C. A., and Chen, Y. 2016. "Revealing Sensitive Information in Personal Interviews: Is Self-Disclosure Easier with Humans or Avatars and under What Conditions?," *Computers in Human Behavior* (65), pp. 23–30.

Pilz, D., and Gewald, H. 2013. "Does Money Matter? Motivational Factors for Participation in Paid- and Non-Profit-Crowdsourcing Communities," in *Proceedings of the 11th International Conference on Wirtschaftsinformatik*, Leipzig, Germany, pp. 73–82.

Podsakoff, P., MacKenzie, S., Lee, J. Y., and Podsakoff, N. P. 2003. "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies," *Journal of Applied Psychology* (88:5), p. 879.

Poetz, M. K., and Schreier, M. 2012. "The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas?," *Journal of Product Innovation Management* (29:2), pp. 245–256.

Pries-Heje, J., Baskerville, R., and Venable, J. 2008. "Strategies for Design Science Research Evaluation," in *16th European Conference on Information Systems, ECIS 2008*, Paper 87.

Provost, F., and Fawcett, T. 2013. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*, " O'Reilly Media, Inc.".

Pumplun, L., Tauchert, C., and Heidt, M. 2019. "A New Organizational Chassis for Artificial Intelligence - Exploring Organizational Readiness Factors," in *Proceedings of the 27th European Conference on Information Systems*, Stockholm, Sweden, June 8.

PwC 2021. "2021 AI Predictions 2021: No Uncertainty Here" (https://www.pwc.com/us/en/services/consulting/library/artificial-intelligence-predictions-2021/no-uncertainty-here.html, accessed May 9, 2021).

Qiu, L., and Benbasat, I. 2008. "Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems," *Journal of Management Information Systems* (25:4), pp. 145–182.

Qureshi, I., and Compeau, D. 2009. “Assessing Between-Group Differences in Information Systems Research: A Comparison of Covariance- and Component-Based SEM,” *MIS Quarterly* (33:1), pp. 197–214.

Radel, R., Pelletier, L. G., Sarrazin, P., and Milyavskaya, M. 2011. “Restoration Process of the Need for Autonomy: The Early Alarm Stage,” *Journal of Personality and Social Psychology* (101:5), pp. 919–934.

Rai, A., Constantinides, P., and Sarker, S. 2019. “Next Generation Digital Platforms: Toward Human-AI Hybrids,” *Mis Quarterly* (43:1), iii–ix.

Rai, A., Pavlou, P., Im, G., and Du, S. 2012. “Interfirm IT Capability Profiles and Communications for Cocreating Relational Value: Evidence from the Logistics Industry,” *MIS Quarterly* (36:1), pp. 233–262.

Rana, R., Staron, M., Hansson, J., Nilsson, M., and Meding, W. 2014. “A Framework for Adoption of Machine Learning in Industry for Software Defect Prediction,” in *Proceedings of 9th International Conference on Software Engineering and Applications (ICSOFT-EA)*, Vienna, Austria, pp. 383–392.

Ransbotham, S., Kiron, D., Gerbert, P., and Reeves, M. 2017. “Reshaping Business With Artificial Intelligence,” *MIT Sloan Management Review* (59:1).

Rawal, K. 2019. “Servitization – The Future” (https://www.tcs.com/blogs/servitization-the-future, accessed April 26, 2019).

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. 2016. “You Only Look Once: Unified, Real-Time Object Detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788.

Reese, H. 2016. “Top 10 AI Failures Of” (https://www.techrepublic.com/article/top-10-ai-failures-of-2016/, accessed November 20, 2018).

Riasanow, T., Flötgen, R. J., Soto Setzke, D., Böhm, M., and Krcmar, H. 2018. “The Generic Ecosystem and Innovation Patterns of the Digital Transformation in the Financial Industry,” in *22nd Pacific Asia Conference on Information Systems (PACIS)*, Yokohama, Japan, Paper 77.

Ribeiro, M. T., Singh, S., and Guestrin, C. 2016. “‘Why Should I Trust You?,’” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’16*, New York, NY, USA, pp. 1135–1144.

Ringle, C. M., Wende, S., and Becker, J.-M. 2015. *SmartPLS 3*, Hamburg: SmartPLS GmbH.

Robles-Flores, J. A., and Roussinov, D. 2012. “Examining Question-Answering Technology from the Task Technology Fit Perspective,” *Communications of the Association for Information Systems* (30), pp. 439–454.

Rogers, E. M. 1995. *Diffusion of Innovations*, (4th ed.), New York, USA: Free Press.

Rogers, E. M. 2003. *Diffusion of Innovations*, (5th ed.), New York, USA: Free Press.

Rose, S., Engel, D., Cramer, N., and Cowley, W. 2010. “Automatic Keyword Extraction from Individual Documents,” in *Text Mining*, Chichester, UK: John Wiley & Sons, Ltd, pp. 1–20.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., and Camerer, C. 1998. “Not so Different after All: A Cross-Discipline View of Trust,” *The Academy of Management Review* (23:3), pp. 393–404.

Rousseeuw, P. J. 1987. “Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis,” *Journal of Computational and Applied Mathematics* (20), pp. 53–65.

Russel, S. J., and Norvig, P. 2009. *Artificial Intelligence: A Modern Approach*, (3rd ed.), Upper Saddle River, NJ, USA: Prentice Hall Press.

Ryan, R. M., and Deci, E. L. 2000. “Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions,” *Contemporary Educational Psychology* (25:1), pp. 54–67.

Rzepka, C., and Berger, B. 2018. “User Interaction with AI-Enabled Systems : A Systematic Review of IS Research,” in *Thirty Ninth International Conference on Information Systems*, San Francisco, pp. 1–17.

Sah, S., Moore, D. A., and MacCoun, R. J. 2013. “Cheap Talk and Credibility: The Consequences of Confidence and Accuracy on Advisor Credibility and Persuasiveness,” *Organizational Behavior and Human Decision Processes* (121:2), pp. 246–255.

Saldaña, J. 2009. *The Coding Manual for Qualitative Researchers,* (1st ed.), London, UK: Sage.

Saldaña, J. 2015. *The Coding Manual for Qualitative Researchers*, (2nd ed.), London, UK: Sage.

Samek, W., Wiegand, T., and Müller, K. R. 2017. “Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models,” *ITU Journal: ICT Discoveries* (1:1), pp. 39–48.

Sarker, S., and Valacich, J. S. 2010. “An Alternative to Methodological Individualism: A Non-Reductionist Approach to Studying Technology Adoption by Groups,” *MIS Quarterly* (34:4), pp. 779–808.

Sarker, S., Xiao, X., and Beaulieu, T. 2013. “Qualitative Studies in Information Systems: A Criti-Cal Review and Some Guiding Principles,” *MIS Quarterly* (37:4), ii–xviii.

Schryen, G., Wagner, G., and Benlian, A. 2015. “Theory of Knowledge for Literature Reviews : An Epistemological Model , Taxonomy and Empirical Analysis of IS Literature,” in *Thirty Sixth International Conference on Information Systems*, Fort Worth, USA, December 13, pp. 1–22.

Schuetzler, R. M., Grimes, G. M., Giboney, J. S., and Nunamaker, J. F. 2018. "The Influence of Conversational Agents on Socially Desirable Responding," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, pp. 283–292.

Schultze, T., Rakotoarisoa, A.-F., and Schulz-Hardt, S. 2015. "Effects of Distance between Initial Estimates and Advice on Advice Utilization," *Judgment and Decision Making* (10:2), pp. 144–171.

Schwartz, R., Dodge, J., Smith, N. A., and Etzioni, O. 2020. "Green AI," *Communications of the ACM* (63:12), pp. 54–63.

Shao, B., Shi, L., Xu, B., and Liu, L. 2012. "Factors Affecting Participation of Solvers in Crowdsourcing: An Empirical Study from China," *Electronic Markets* (22:2), pp. 73–82.

Sheeran, P. 2002. "Intention—Behavior Relations: A Conceptual and Empirical Review," *European Review of Social Psychology* (12:1), pp. 1–36.

Siddiqi, S., and Sharan, A. 2015. "Keyword and Keyphrase Extraction Techniques: A Literature Review," *International Journal of Computer Applications* (109:2), pp. 18–23.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., and Lanctot, M. 2016. "Mastering the Game of Go with Deep Neural Networks and Tree Search," *Nature* (529:7587), pp. 484–489.

Simon, H. A. 1972. "Theories of Bounded Rationality," *Decision and Organization* (1:1), pp. 161–176.

Sironi, P. 2016. "The Theory of Innovation: From Robo-Advisors to Goal Based Investing and Gamification," *John Wiley & Sons*.

Skitka, L. J., Mosier, K. L., and Burdick, M. 1999. "Does Automation Bias Decision-Making?," *International Journal of Human Computer Studies* (51:5), pp. 991–1006.

Sniezek, J. A., and Buckley, T. 1995. "Cueing and Cognitive Conflict in Judge-Advisor Decision Making," *Organizational Behavior and Human Decision Processes* (62:2), pp. 159–174.

Sniezek, J. A., and Van Swol, L. M. 2001. "Trust, Confidence, and Expertise in a Judge-Advisor System," *Organizational Behavior and Human Decision Processes* (84:2), pp. 288–307.

Statista 2019. "Robo-Advisors - Worldwide" (https://www.statista.com/outlook/337/100/robo-advisors/worldwide, accessed March 1, 2019).

Statista 2020. "Volume of Data/Information Created, Captured, Copied, and Consumed Worldwide from 2010 to 2024" (https://www.statista.com/statistics/871513/worldwide-data-created/, accessed May 13, 2021).

Straub, D., and del Giudice, M. 2012. “Editor’s Comments: Use,” *MIS Quarterly* (36:4), iii–vii.

Strickland, E. 2019. “How IBM Watson Overpromised and Underdelivered on AI Health Care - IEEE Spectrum,” *IEEE Spectrum* (https://spectrum.ieee.org/biomedical/diagnostics/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care, accessed May 13, 2021).

Sturm, T., Gerlach, J., Pumplun, L., Mesbah, N., Peters, F., Tauchert, C., Nan, N., and Buxmann, P. 2021. “Coordinating Human and Machine Learning for Effective Learning,” *MIS Quarterly* (Forthcoming).

Van Swol, L. M. 2011. “Forecasting Another’s Enjoyment versus Giving the Right Answer: Trust, Shared Values, Task Effects, and Confidence in Improving the Acceptance of Advice,” *International Journal of Forecasting* (27:1), pp. 103–120.

Van Swol, L. M., and Sniezek, J. A. 2005. “Factors Affecting the Acceptance of Expert Advice,” *British Journal of Social Psychology* (44:3), pp. 443–461.

Tauchert, C., Bender, M., Mesbah, N., and Buxmann, P. 2020. “Towards an Integrative Approach for Automated Literature Reviews Using Machine Learning,” in *Proceedings of the 53rd Hawaii International Conference on System Sciences*, Weilea, Hawaii, USA, January 7.

Tauchert, C., Buxmann, P., and Lambinus, J. 2020. “Crowdsourcing Data Science: A Qualitative Analysis of Organizations’ Usage of Kaggle Competitions,” in *Proceedings of the 53rd Hawaii International Conference on System Sciences*, Weilea, Hawaii, USA, January 7.

Tauchert, C., and Mesbah, N. 2019. “Following the Robot? Investigating Users’ Utilization of Advice from Robo-Advisors,” in *Proceedings of the 40th International Conference on Information Systems*, Munich, Germany, December 18.

Tertilt, M., and Scholz, P. 2017. “To Advise, or Not to Advise — How Robo-Advisors Evaluate the Risk Preferences of Private Investors,” *The Journal of Wealth Management* (21:2), pp. 70–84.

The Economist 2017. “Regulating the Internet Giants - The World’s Most Valuable Resource Is No Longer Oil, but Data” (https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data, accessed November 3, 2018).

Thorndike, R. L. 1953. “Who Belongs in the Family?,” *Psychometrika* (18:4), pp. 267–276.

Tintarev, N., and Masthoff, J. 2007. “A Survey of Explanations in Recommender Systems,” in *IEEE 23rd International Conference on Data Engineering Workshop*, pp. 801–810.

Trentin, A., Perin, E., and Forza, C. 2012. “Product Configurator Impact on Product Quality,” *International Journal of Production Economics* (135:2), pp. 850–859.

Tsichritzis, D. 1997. *The Dynamics of Innovation BT - Beyond Calculation: The Next Fifty Years of Computing*, P. J. Denning and R. M. Metcalfe (eds.), New York, NY: Springer New York, pp. 259–265.

United Nations 2008. *International Standard Industrial Classification of All Economic Activities*, (4th ed.), New York, USA: United Nations Publication.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. 2003. “User Acceptance of Information Technology: Toward a Unified View,” *MIS Quarterly* (27:3), p. 425.

Venkatesh, V., Thong, J. Y. L., and Xu, X. 2012. “Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology,” *MIS Quarterly*, pp. 157–178.

de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., and Parasuraman, R. 2016. “Almost Human: Anthropomorphism Increases Trust Resilience in Cognitive Agents,” *Journal of Experimental Psychology: Applied* (22:3), pp. 331–349.

Wærn, Y., and Ramberg, R. 1996. “People’s Perception of Human and Computer Advice,” *Computers in Human Behavior* (12:1), pp. 17–27.

Wang, C., and Blei, D. M. 2011. “Collaborative Topic Modeling for Recommending Scientific Articles,” in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’11*, New York, New York, USA, p. 448.

Wang, W., and Benbasat, I. 2005. “Trust in and Adoption of Online Recommendation Agents,” *Journal of the Association for Information Systems* (6:3), Article 4.

Wang, W., and Benbasat, I. 2007. “Recommendation Agents for Electronic Commerce: Effects of Explanation Facilities on Trusting Beliefs,” *Journal of Management Information Systems* (23:4), pp. 217–246.

Wanner, J., Heinrich, K., Janiesch, C., and Zschech, P. 2020. “How Much AI Do You Require? Decision Factors for Adopting AI Technology,” in *Proceedings of the 41st International Conference on Information Systems*, India, December 14.

Weber, R. P. 1990. *Basic Content Analysis*, (2nd ed.), Newbury Park, CA: Sage.

Webster, J., and Watson, R. T. 2002. “Analyzing the Past to Prepare for the Future: Writing a Literature Review,” *MIS Quarterly* (26:2), xiii–xxiii.

Weinfurt, K. 2002. “Repeated Measures Analyses: ANOVA MANOVA, and HLM,” *Reading and Understanding More Multivariate Statistics*, pp. 317–361.

White, T. B. 2005. “Consumer Trust and Advice Acceptance: The Moderating Roles of Benevolence, Expertise, and Negative Emotions,” *Journal of Consumer Psychology* (15:2), pp. 141–148.

Winnefeld, C. H., and Permantier, A. 2017. “FinTech - The Digital (R)Evolution in the German Banking Sector?,” *Business and Management Research* (6:3), pp. 65–84.

Xu, H., Hock-Hai, T., Tan, B. C. Y., and Agarwal, R. 2012. “Effects of Individual Self-Protection, Industry Self-Regulation, and Government Regulation on Privacy Concerns: A Study of Location-Based Services,” *Information Systems Research* (23:4), pp. 1342–1363.

Xu, J., Benbasat, I., and Cenfetelli, R. T. 2014. “The Nature and Consequences of Trade-off Transparency in the Context of Recommendation Agents,” *MIS Quarterly* (38:2), pp. 379–406.

Yan, D., Zhou, Q., Wang, J., and Zhang, N. 2016. “Bayesian Regularisation Neural Network Based on Artificial Intelligence Optimisation,” *International Journal of Production Research* (55:8), pp. 2266–2287.

Yang, J., Adamic, L. A., and Ackerman, M. S. 2008. “Crowdsourcing and Knowledge Sharing: Strategic User Behavior on Taskcn,” in *Proceedings of the 9th ACM Conference on Electronic Commerce*, pp. 246–255.

Yaniv, I. 2004. “Receiving Other People’s Advice: Influence and Benefit,” *Organizational Behavior and Human Decision Processes* (93:1), pp. 1–13.

Ye, L. R., and Johnson, P. E. 1995. “The Impact of Explanation Facilities on User Acceptance of Expert Systems Advice,” *MIS Quarterly* (19:2), p. 157.

Yin, R. K. 2009. *Case Study Research: Design and Methods*, (4th ed.), Los Angeles, USA: SAGE Publications.

Yoav Shoham, Raymond Perrault, Erik Brynjolfsson, Jack Clark, James Manyika, Juan Carlos Niebles, Terah Lyons, John Etchemendy, Barbara Grosz, and Zoe Bauer 2018. “The AI Index 2018 Annual Report,” *AI Index Steering Committee, Human-Centered AI Initiative, Stanford University, Stanford, CA* (https://hai.stanford.edu/ai-index-2018, accessed May 19, 2021).

Zanker, M. 2012. “The Influence of Knowledgeable Explanations on Users’ Perception of a Recommender System,” in *Proceedings of the Sixth ACM Conference on Recommender Systems*, pp. 269–272.

Zhao, Y., and Zhu, Q. 2014. “Evaluation on Crowdsourcing Research: Current Status and Future Direction,” *Information Systems Frontiers* (16:3), pp. 417–434.

Zheng, H., Li, D., and Hou, W. 2011. "Task Design, Motivation, and Participation in Crowdsourcing Contests," *International Journal of Electronic Commerce* (15:4), pp. 57–88.

Zhu, K., and Kraemer, K. L. 2005. "Post-Adoption Variations in Usage and Value of E-Business by Organizations: Cross-Country Evidence from the Retail Industry," *Information Systems Research* (16:1), pp. 61–84.

Zigurs, I., and Buckland, B. K. 1998. "A Theory of Task/Technology Fit and Group Support Systems Effectiveness," *MIS Quarterly* (2:3), pp. 313–334.

Zigurs, I., Buckland, B. K., Connolly, J. R., and Wilson, E. V. 1999. "A Test of Task-Technology Fit Theory for Group Support Systems," *ACM SIGMIS Database* (30:3–4), pp. 34–50.

Zuboff, S. 1988. *In the Age of the Smart Machine: The Future of Work and Power*, New York, NY, USA: Basic Books.

Appendix

A1. Constructs' Items (Paper C)

RAL	I would find it more convenient to educate myself about financial assets with the help of a robo-advisor by interacting with it online than by asking questions of a financial expert.
	I would learn more if I was informed about financial assets with the help of a robo-advisor than by talking to a financial expert.
	I would have greater confidence in the explanations provided by a robo-advisor than those offered by a financial expert.
	I would understand the explanations offered by a robo-advisor better than those provided by a financial expert.
RAIT	I believe such a robo-advisor would provide more objective recommendations than a financial expert.
	I would trust the recommendation of such robo-advisor more than the recommendation of a financial expert with regard to the appropriate level of coverage for my needs.
	I would expect a greater return on investment using a robo-advisor than through a financial expert.
	I would trust the accuracy of financial information provided by a robo-advisor more than those provided by a financial expert.
RAIC	I would find it more convenient to use a robo-advisor rather than a financial expert.
	It would be more convenient for me to use a robo-advisor to evaluate financial assets than a financial expert.
RAT	I would find it more convenient to manage financial assets on the Internet through a robo-advisor than through a financial expert.
	I would feel more confident managing financial assets on the Internet through a robo-advisor than through a financial expert.
	I would be confident to assess a financial asset on the Internet through a robo-advisor than through a financial expert.
	I would find it more convenient to assess a financial asset on the Internet through a robo-advisor than through a financial expert.
LOG	If the robo-advisor would enter into a dialogue with me like a human being, my trust would increase.
RES	If the robo-advisor would tell me the most important reasons that led to the recommendation, my trust would increase.
DAT	Detailed information about data which the robo-advisor uses to generate the advice would strengthen my trust in the robo-advisor.
VIS	If the robo-advisor would have a visual appearance, such as a figure, then my trust would increase.
CON	The information how confident the robo-advisor is with his recommendation would strengthen my trust.
HIS	Documentation of the previous recommendations and its return of investment of the robo-advisor would strengthen my trust.
TST	The possibility to work with the robo-advisor first without risk to test it would strengthen my trust.
SOC	Recommendations from friends/acquaintances to use robo-advisors would strengthen my trust.
USE	The information on how long the robo-advisor has been in use would strengthen my trust.
FNC	Information about the technical functionality of the robo-advisors would strengthen my trust.
FRQ	The information how often the robo-advisor is trained / learns would strengthen my trust.

Table 11. Constructs‘ Items

A2. Survey Items (Paper D)

Construct	Item		Adapted from…
Task-Advisor Fit	TAF1	The expert's[2] advisory service is compatible with all aspects of this task.	(Moore and Benbasat 1991)
	TAF2	The expert's advisory service fits very well with my needs in the task.	
	TAF3	The expert's advisory service fits into my way of decision-making.	
Advisor Expertise	AEX1	The expert is competent and effective in estimating the stock price.	(McKnight et al. 2002)
	AEX2	The expert performs its role of estimation the stock price very well.	
	AEX3	Overall, the expert is a capable and proficient advisor for estimating the stock price.	
	AEX4	In general, the expert is very knowledgeable about the stock price prediction.	
Efficiency-Enhancing	EFF	The expert increases the efficiency of my decision making.	(Chan et al. 1997)
Emotional Trust in Advisor	EMO1	I feel secure about relying on the expert for my decision.	(Komiak and Benbasat 2006)
	EMO2	I feel comfortable about relying on the expert for my decision.	
	EMO3	I feel content about relying on the expert for my decision.	
Advisor Integrity	INT1	The expert provides unbiased recommendations.	(Komiak and Benbasat 2006)
	INT2	The expert is honest.	
	INT3	I consider the expert to be of integrity.	
User's Expertise	UEX1	I feel very competent in the above explained task.	(Radel et al. 2011)
	UEX2	I feel able to meet the challenge of performing well in this task.	
	UEX3	I am able to master this task.	
	UEX4	I am good at doing this task.	

Table 12. Survey Items

[2] Depending on the experimental group, the term "expert" is replaced by "human expert" or "robo-advisor" in all items.

A3. Results Visualization (Paper E)

Figure 16. Dendrogram of All Publications with Cluster Titles

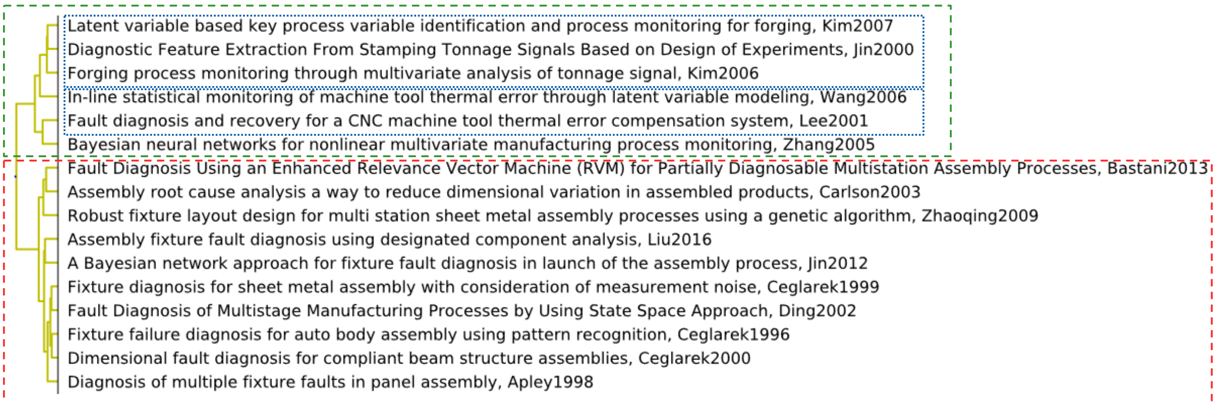

Figure 17. Dendrogram with Titles and Authors for Cluster 1

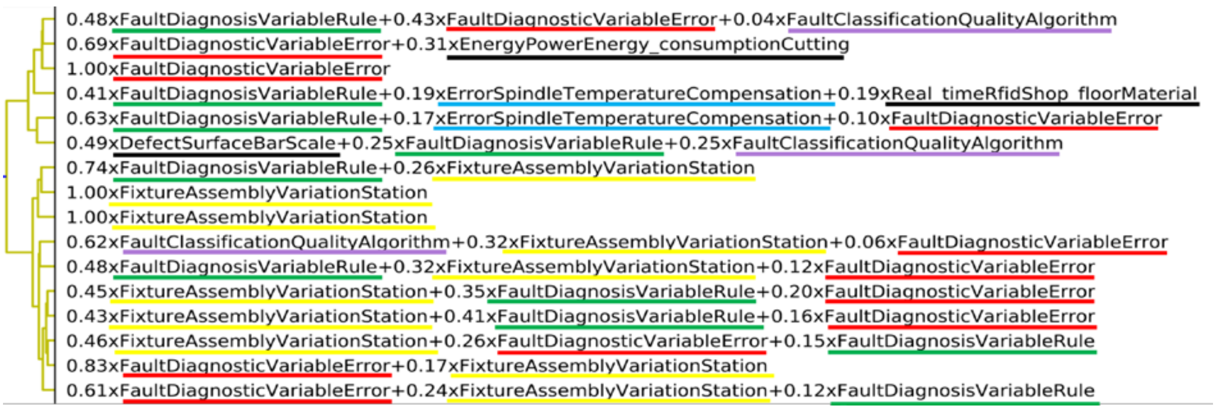

Figure 18. Dendrogram with Topics for Cluster 1