Following the Robot – Investigating the Utilization and the Acceptance of AI-based Services

Department of Law and Economics
at the Technical University of Darmstadt

Doctoral thesis

by Neda Mesbah

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

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

Neda Mesbah

Following the Robot – Investigating the Utilization and the Acceptance of AI-based Services

Darmstadt, Germany, Technical University of Darmstadt

Dissertation published on TUprints in 2023

Date of the viva voce: 07/14/2023

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.

Neda Mesbah

Darmstadt, 5th April 2023

Abstract

In the past few years, there has been significant progress in the field of artificial intelligence (AI), with advancements in areas such as natural language processing and machine learning. AI systems are now being used in various industries and applications, from healthcare to finance, and are becoming more sophisticated and capable of handling complex tasks. The technology has the potential to assist in both private and professional decision-making. However, there are still challenges to be addressed, such as ensuring transparency and accountability in AI decision-making processes and addressing issues related to bias and ethics, and it is not yet certain whether all of these newly developed AI-based services will be accepted and used.

This thesis addresses a research gap in the field of AI-based services by exploring the acceptance and utilization of such services from both individual and organizational perspectives. The research examines various factors that influence the acceptance of AI-based services and investigates users' perceptions of these services. The thesis poses four research questions, including identifying the differences in utilizing AI-based services compared to human-based services for decision-making, identifying characteristics of acceptance and utilization across different user groups, prioritizing methods for promoting trust in AI-based services, and exploring the impact of AI-based services on an organization's knowledge.

To achieve this, the study employs various research methods such as surveys, experiments, interviews, and simulations within five research papers. Research focused on an organization that offers robo-advisors as an AI-based service, specifically a financial robo-advisor. This research paper measures advice-taking behavior in the interaction with robo-advisors based on the judge-advisor system and task-technology fit frameworks. The results show the advice of robo-advisors is followed more than that of human advisors and this behavior is reflected in the task-advisor fit. Interestingly, the advisor's perceived expertise is the most influential factor in the task-advisor fit for both robo-advisors and human advisors. However, integrity is only significant for human advisors, while the user's perception of the ability to make decisions efficiently is only significant for robo-advisors.

Research paper B examined the differences in advice utilization between AI-based and human advisors and explored the relationship between task, advisor, and advice utilization using the task-advisor fit just like research paper A but in context the of a guessing game. The research paper analyzed the impact of advice similarity on utilization. The results indicated that judges tend to use advice from AI-based advisors more than human advisors when the advice is similar to their own estimation. When the advice is vastly different from their estimation, the utilization rate is equal for both AI-based and human advisors.

Research paper C investigated the different needs of user groups in the context of health chatbots. The increasing number of aging individuals who require considerable medical attention could be addressed by health chatbots capable of identifying diseases based on symptoms. Existing chatbot applications are primarily used by younger generations. This research paper investigated the factors affecting the adoption of health chatbots by older people and the extended Unified Theory of Acceptance and Use of Technology.

To investigate how to promote AI-based services such as robo-advisors, research paper D evaluated the effectiveness of eleven measures to increase trust in AI-based advisory systems and found that noncommittal testing was the most effective while implementing human traits had negligible effects. Additionally, the relative advantage of AI-based advising over that of human experts was measured in the context of financial planning. The results suggest that convenience is the most important advantage perceived by users.

To analyze the impact of AI-based services on an organization's knowledge state, research paper E explored how organizations can effectively coordinate human and machine learning (ML). The results showed that ML can decrease an organization's need for humans' explorative learning. The findings demonstrated that adjustments made by humans to ML systems are often beneficial but can become harmful under certain conditions. Additionally, relying on knowledge created by ML systems can facilitate organizational learning in turbulent environments, but it requires significant initial setup and coordination with humans. These findings offer new perspectives on organizational learning with ML and can guide organizations in optimizing resources for effective learning.

In summary, the findings suggest that the acceptance and utilization of AI-based services can be influenced by the fit between the task and the service. However, organizations must carefully consider the user market and prioritize mechanisms to increase acceptance. Additionally, the implementation of AI-based services can positively affect an organization's ability to choose learning strategies or navigate turbulent environments, but it is crucial for humans to maintain domain knowledge of the task to reconfigure such services. This thesis enhances our understanding of the acceptance and utilization of AI-based services and provides valuable insights on how organizations can increase customers' acceptance and usage of their AI-based services as well as implement and use AI-based services effectively.

Abstract (German version)

In den letzten Jahren hat es auf dem Gebiet der künstlichen Intelligenz (KI) erhebliche Fortschritte gegeben, wie z.B. im Bereich der Verarbeitung natürlicher Sprache oder beim maschinellen Lernen. KI-Systeme werden inzwischen in verschiedenen Branchen und Anwendungen eingesetzt, vom Gesundheitswesen bis zum Finanzwesen. Zusätzlich werden diese immer ausgefeilter und können immer komplexere Aufgaben bewältigen. Die Technologie hat das Potenzial, sowohl im privaten als auch im beruflichen Bereich bei der Entscheidungsfindung zu unterstützen. Allerdings gibt es auch noch einige Herausforderungen zu bewältigen. Dies betrifft etwa die Gewährleistung von Transparenz und die Frage nach Haftung bei KI-Entscheidungsprozessen. Darüber hinaus gibt es einige Unsicherheiten im Zusammenhang mit Themen wie Verzerrung und Ethik, daher ist es noch nicht sicher, ob solche Dienste eine Breite Akzeptanz finden und genutzt werden.

Die vorliegende Arbeit befasst sich mit dieser Forschungslücke im Bereich der KI-basierten Dienste, indem sie die Akzeptanz und Nutzung solcher Dienste sowohl aus individueller als auch aus organisatorischer Sicht untersucht. Die Arbeit untersucht dabei verschiedene Faktoren, die die Akzeptanz von KI-basierten Diensten beeinflussen, und untersucht die Wahrnehmung dieser Dienste durch die Nutzer[1]. Die Dissertation stellt hierzu vier Forschungsfragen auf, darunter die Identifizierung der Unterschiede bei der Nutzung von KI-basierten Diensten im Vergleich zu menschenbasierten Diensten, die Identifizierung von Akzeptanz- und Nutzungskritierien in verschiedenen Nutzergruppen, die Priorisierung von Methoden zur Förderung des Vertrauens in KI-basierte Dienste und die Untersuchung der Auswirkungen von KI-basierten Diensten auf den Wissensstand einer Organisation.

Um dies zu erreichen, werden verschiedene Forschungsmethoden wie Umfragen, Experimente, Interviews und Simulation in fünf Forschungspapieren eingesetzt. Forschungspapier A konzentriert sich auf Organisationen, die Robo-Advisors, insbesondere Finanz-Robo-Advisors, als KI-basierte Dienstleistung anbieten. Im Rahmen dieses Forschungspapiers wird das tatsächliche Beratungsverhalten in der Interaktion mit Robo-Advisors auf der Grundlage von Judge-Advisor-Systemen und Task-Technology-Fit-Frameworks gemessen. Die Ergebnisse zeigen, dass die

[1] Im Folgenden wird aus Gründen der besseren Lesbarkeit ausschließlich die männliche Form verwendet. Sie bezieht sich auf Personen beiderlei Geschlechts.

Ratschläge von Robo-Beratern eher befolgt werden als die von menschlichen Beratern. Dieses Verhalten spiegelt sich im Task-Advisor-Fit wider. Interessanterweise ist die wahrgenommene Expertise des Beraters der einflussreichste Faktor auf den Task-Advisor-Fit sowohl für Robo-Advisors als auch für Human-Advisors. Die Integrität ist jedoch nur bei menschlichen Beratern signifikant, während die vom Nutzer wahrgenommene Fähigkeit, Entscheidungen effizient zu treffen, nur bei Robo-Advisors signifikant ist.

Forschungsarbeit B untersucht die Unterschiede in der Beratungsnutzung zwischen KI-basierten und menschlichen Beratern und untersucht die Beziehung zwischen Aufgabe, Berater und Beratungsnutzung unter Verwendung des Task-Technology-Fit. Dieses Vorgehen ist ähnlich zu Forschungsarbeit A, den Kontext bilden hier jedoch Ratespiele. Darüber hinaus wurde der Einfluss der Ähnlichkeit der Ratschläge auf die Nutzung analysiert. Die Ergebnisse zeigen, dass Richter dazu neigen, Ratschläge von KI-basierten Beratern mehr zu nutzen als menschliche Berater, wenn die Ratschläge ihrer eigenen Einschätzung ähneln. Weicht der Ratschlag stark von ihrer Einschätzung ab, ist die Nutzungsrate bei KI-gestützten und menschlichen Beratern gleich hoch.

Forschungspapier C untersucht die unterschiedlichen Bedürfnisse von Nutzergruppen im Kontext von Gesundheits-Chatbots. Die zunehmende Anzahl älterer Menschen, die eine hohe medizinische Fürsorge benötigen, könnte durch Gesundheits-Chatbots, die in der Lage sind, Krankheiten anhand von Symptomen zu erkennen, adressiert werden. Bestehende Chatbot-Anwendungen werden hauptsächlich von jüngeren Generationen genutzt. Dieses Forschungspapier untersucht die Faktoren, die sich auf die Annahme von Gesundheits-Chatbots durch Senioren auswirken, und die erweiterte einheitliche Theorie der Akzeptanz und Nutzung von Technologien.

Um herauszufinden, wie KI-basierte Dienste wie Robo-Advisors gefördert werden können, bewertet das Forschungspapier D die Wirksamkeit von elf Maßnahmen zur Steigerung des Vertrauens in KI-basierte Beratungssysteme. Das Resultat hierbei lautet, dass unverbindliche Tests am effektivsten sind, während die Implementierung menschlicher Eigenschaften vernachlässigbare Auswirkungen haben. Außerdem wurde der relative Vorteil von KI-gestützter Beratung gegenüber menschlichen Experten im Kontext der Finanzplanung gemessen. Die Ergebnisse deuten darauf hin, dass die Bequemlichkeit der wichtigste Vorteil ist, den die Nutzer wahrnehmen.

Um die Auswirkungen von KI-basierten Diensten auf den Wissensstand eines Unternehmens zu analysieren, untersucht Forschungspapier E, wie Unternehmen menschliches und ML-Lernen effektiv koordinieren können. Die Ergebnisse zeigen, dass ML den Bedarf einer Organisation an menschlichem explorativem Lernen verringern kann. Die Ergebnisse zeigen, dass die von Menschen vorgenommenen Anpassungen an ML-Systeme oft nützlich sind, unter bestimmten

Bedingungen aber auch schädlich sein können. Darüber hinaus kann der Rückgriff auf das von ML-Systemen erzeugte Wissen das organisatorische Lernen in turbulenten Umgebungen erleichtern. Dies erfordert jedoch eine umfangreiche anfängliche Einrichtung der Systeme und Koordination mit Menschen. Diese Erkenntnisse bieten neue Perspektiven für das organisatorische Lernen mit ML und können Organisationen bei der Optimierung von Ressourcen für effektives Lernen unterstützen.

Zusammenfassend legen die Ergebnisse nahe, dass die Akzeptanz und Nutzung von KI-basierten Diensten durch das Zusammenpassen zwischen Aufgabe und Dienst beeinflusst werden kann. Unternehmen müssen jedoch den Nutzermarkt sorgfältig berücksichtigen und Mechanismen zur Steigerung der Akzeptanz priorisieren. Darüber hinaus kann sich die Implementierung von KI-basierten Diensten positiv auf die Fähigkeit einer Organisation auswirken, Lernstrategien zu wählen oder sich in turbulenten Umgebungen zurechtzufinden. Es ist aber von entscheidender Bedeutung, dass Menschen das Domänenwissen über die Aufgabe beibehalten, um solche Dienste neu zu konfigurieren. Insgesamt erweitert diese Arbeit das Verständnis für die Akzeptanz und Nutzung von KI-basierten Diensten und liefert wertvolle Erkenntnisse für Unternehmen, um die Akzeptanz und Nutzung ihrer KI-basierten Dienste durch die Kunden zu erhöhen. Zudem liefert sie einen Beitrag darüber, wie man KI-basierte Dienste in Unternehmen implementiert und effektiv nutzt.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AEX	Advisor Expertise
AI	Artificial Intelligence
AMCIS	Americas Conference on Information Systems
App	Application
AVE	Average Variance Extracted
CA	Close Advice
CON	Confidence
CR	Composite Reliability
Cr. α	Cronbach's Alpha
DAT	Data Transparency
ECIS	European Conference on Information Systems
EFF	Advisor Efficient-Enhancing
EMO	Emotional Trust in Advisor
FA	Far Advice
FNC	Technical Functionality
FRQ	Training Frequency
H	Hypothesis
HC	Health Chatbots
HCI	Human-Computer Interaction
HICSS-54	54th Hawaii International Conference on System Sciences
HIS	History
ICIS	International Conference on Information Systems
INT	Advisor Integrity

IS	Information Systems
IT	Information Technology
JAS	Judge-Advisor Systems
LOG	Dialog
M	Mean
mHealth	Mobile Health
ML	Machine Learning
NLP	Natural Language Processing
PLS	Partial Least Square
RoA	Robo-Advisors
RA	Perceived Relative Advantage
RES	Reasoning
RQ	Research Question
SD	Standard Deviation
SOC	Social Environment
TAF	Task-Addvisor Fit
TAM	Technology Acceptance Model
TST	Testing
TTF	Task-Technology Fit
UEX	User’s Self-Perceived Expertise
USE	Usage Time
UTAUT	Unified Theory of Acceptance and Use of Technology
UTAUT2	Extended Unified Theory of Acceptance and Use of Technology
VHB	Verband der Hochschullehrer für Betriebswirtschaft e.V.
VIS	Visual Appearance
WOA	Weight Of Advice

1 Introduction

"Artificial intelligence (AI), manifested by machines that exhibit aspects of human intelligence […], is increasingly utilized in service and today is a major source of innovation […]."

Ming-Hui Hang and Roland T. Rust (Huang and Rust 2018)

Overarching Motivation

An average adult makes many decisions every day. For example, adults make over 200 decisions each day just about food (Wansink and Sobal 2007). Therefore, it can be assumed that numerous decisions are made in private and professional life; it is often estimated that thousands of decisions are made each day (e.g., Lally 2016; Szalavitz 2012). Many of these decisions are not made alone; rather, advice is sought from other people such as parents, friends, or experts. Experts in particular provide important decision-making assistance when personal knowledge or experience is lacking (Sniezek and Van Swol 2001). Nowadays, not only human experts with knowledge and experience exist to support us in decision-making but also machines.

Technology has evolved significantly in the past few years, and in the age of digitalization, many aspects of our lives are influenced by technology and its ability to assist us in the role of experts. Artificial intelligence is a more recent development in which machines act as experts and provide advice in areas that previously only human experts could comprehend. There are numerous definitions of what artificial intelligence means. One common definition describes AI as the "science and engineering of making intelligent machines, especially intelligent computer programs" through a simulation of human intelligence using underlying technologies such as machine learning, deep learning, and natural language processing (Elliot and Andrews 2017; McCarthy 2007, p. 2). Another definition says AI is a machine that has the ability to learn, makes rational predictions, and interacts like a human (Russell and Norvig 2016). The fundamental similarity is that they consider an AI to be an intelligent system that is similar to humans. The trends and developments in artificial intelligence have become particularly interesting in recent years due to technical developments in hardware that allow better implementation of machine learning methods, especially neural networks, and thus offer more opportunities to perform tasks or services with machines. Such machines accordingly have a great potential to support us in decision-making in our private and professional lives, such that we see significant changes to our everyday activities with the use of such machines and can expect even more substantial changes (Thomas and Powers 2022).

AI-based services have advantages and disadvantages. One of the disadvantages is that the recommendations of AI-based services are often not as explainable as those of humans, which is known as black-box behavior (Cassauwers 2020). However,due to technological advances, AI-based algorithms can process, utilize, and learn from more information than any human could in a comparable time frame because of their cognitive constraints (Simon 1972). Moreover, AI is permanently available, its performance does not decrease over time, and it is cheaper (Tertilt and Scholz 2017). Two of the main advantages of AI-based services are efficiency and scalability (Brundage et al. 2018).

There are already numerous AI-based services available to support us. In the private sector, there are several examples that range from intelligent toothbrushes that use AI to determine which areas of the mouth have not been brushed sufficiently (Meisenzahl 2019), to financial robo-advisors that enable an automated investment advisory process (Jung, Dorner, Glaser, et al. 2018). Thus, such services concern not only low-stakes decisions, such as gaming, but also increasingly high-stakes decisions, such as finances and health. At work, we are also constantly coming into contact with various AI-based services, such as intelligent knowledge management tools, which enable a smoother and faster onboarding process for employees (Starmind 2021) or with market segmentation tools, which classify current and potential customers with regard to their industry segment and functional role (Lieder et al. 2019). Hence, even in the world of work, tasks of both high and low importance can be solved with the help of AI-based services.

These examples show that there are many possible touch points with AI-based services, both in private and professional contexts. However, it is not certain whether these AI-based services will be accepted and used. In previous research, various approaches have investigated the acceptance of these technologies. There are varied models for this purpose, such as the technology acceptance model (TAM) (Davis 1989) or the unified theory of acceptance and use of technology (UTAUT) (Venkatesh et al. 2003). The vast majority of these approaches have assumed that IT is only a tool that supports human learning (e.g., Gefen et al. 2003; Lee 2009; Lu and Yang 2014). However, this assumption neglects AI-based services' ability to learn autonomously and contribute their self-learned knowledge (e.g., Ransbotham et al. 2020; Seidel et al. 2019). Therefore, AI differs significantly from other traditional technologies because AI-based services do not just follow predefined static rules but have the ability to learn from data (Burrell 2016). In the cognitive sciences, the judge-advisor system (JAS) paradigm has been used to investigate the behavior of people giving and taking advice. Although various factors were examined within this research stream, it was almost exclusively focused on the interaction between human decision-makers and human advisors (Bonaccio and Dalal 2006). Thus, specific prior research about the acceptance and utilization of technology and advice has to be proven and extended for AI-based services. Given the rapid development of technology and the constantly increasing amount of AI-based services, it seems increasingly important to further explore the acceptance and utilization of AI-based services. Therefore, this thesis examines whether AI-based services can provide reasonable assistance in decision-making to raise the level of acceptance and utilization of such services so that people can benefit from this technology.

The findings of this thesis contribute to the information system and cognitive science by enabling a better understanding of the acceptance and utilization of AI-based services. Moreover, the providers of such AI-based services can use the findings to predict whether users would accept and use their product and how to promote its use. In addition, organizations can use the findings of this thesis to discover how the utilization of such AI-based services affects the knowledge level of their organization.

Overarching Research Questions and Contribution

Due to a lack of knowledge about the acceptance and utilization of AI-based services, this thesis addresses this research gap from an individual and organizational point of view. AI-based services usually replace services that were previously offered and executed by humans. Whether such a replacement is accepted by an individual depends on many different factors. From prior research on technology acceptance, it is known that one of the most decisive factors is the AI's performance (Venkatesh et al. 2003). Even if it is assumed that the AI's performance is as good as that of a human, it is not assured that such AI-based services will be used. Therefore, it should be determined if and when AI-based services will be accepted and used. But there is not enough research to fully grasp this issue. Researchers often focus on the design and architecture of such services and related business models (Eickhoff et al. 2017; Jung, Dorner, Weinhardt, et al. 2018; Jung and Weinhardt 2018; Riasanow et al. 2018). The exploration of users' perception of AI-based services has been mostly neglected. Recently, there has been more research on this topic but different findings (Jussupow et al. 2020). Some of the findings show that AI-based services are accepted, used, and preferred compared to human services (Logg et al. 2019), whereas other findings show results contradicting these findings (e.g., Castelo et al. 2019; Lee 2018). Therefore, we examine if users will accept the substitution of humans for AI-based services and if these services will be at least similarly utilized. This leads to the following research questions (RQ):

RQ1: Are there differences in users' utilization of AI-based services and human-based services in the decision-making process?

The decision-making process is influenced by many factors, which can be roughly divided into three areas. In particular, decisions are not based only on perceptions of the expertise of an AI or human expert but on the decision maker and other factors that affect the problem itself (e.g., Bonaccio and Dalal 2006; Gino and Moore 2007). The difficulty of the problem is also crucial (Gino and Moore 2007). Accordingly, decisions can be divided between low-stakes decisions, such as gaming, and high-stakes decisions, such as finance. It cannot be assumed that acceptance and utilization of AI-based services for low-stakes decisions are the same as those for high-stakes decisions (Wanner et al. 2020). That is why the first research question should be examined from two perspectives, low-stakes and high-stakes decisions.

Furthermore, the role of the decision-maker should not be neglected. Different user groups have different needs for a product or service and this can affect their perception of the product or service. Thus, the acceptance and utilization of AI-based services may vary depending on the user group. There are numerous user groups and numerous methods used to divide user groups. One classification from the diffusion of innovations theory distinguishes between five user groups: innovators, early adopters, early majority, late majority, and laggards (Rogers 2003). According to diffusion theory, the diffusion of an innovation depends on the user group, and the categorization defines the extent to which an individual accepts a new idea. Thus, innovations diffuse faster among early adopters than among the early majority (Rogers 2003). The acceptance of AI-based services will also be different for the various user groups. There will be early adopters, but there are factors other than innovativeness which distinguish the user groups of AI-based services. This leads to the second research question:

RQ2: How do the acceptance and utilization of AI-based services differ among the different user groups?

Various factors can be influenced to increase the acceptance and utilization of AI-based services for all user groups. One of the key factors in innovation adoption is users' trust in such services (Lin 2011; Pavlou 2018) and it is a key factor in the JAS (Sniezek and Van Swol 2001; Van Swol 2011). It is likely that trust will also play a major role in the acceptance and utilization of AI-based services. Prior research has investigated various opportunities to increase trust in experts or recommender systems, but these studies have usually investigated only one method or mechanism and its effect, such as transparency (Nilashi et al. 2016) or anthropomorphism (de Visser et al. 2016). However, organizations cannot implement all existing practices due to time, financial, and technological constraints. Therefore, the available options should be evaluated and prioritized so that the most effective ones can be implemented. This leads to the third research question:

RQ3: How can AI-based services be promoted to increase trust in such services?

For organizations, it is not only relevant whether AI-based services provide benefits for an individual or whether the individual accepts and uses the services, but also whether the organization in which the individual works can benefit from the service. The utilization of AI-based services has an impact on the knowledge state of an organization, because AI-based services are generating new knowledge which humans cannot (easily) generate (e.g., Choudhury et al. 2021; Ransbotham et al. 2020). Thus, the utilization of such AI-based services within an organization influences organizational learning. Organizations should evaluate whether using AI-based services within their organization has a positive or possibly negative impact on their knowledge state before they promote these services. This leads to the fourth research question:

RQ4: What effect does the utilization of AI-based services have on organizational learning?

Structure of the Thesis

Based on these four research questions, the paper includes five research papers (four focusing on individuals and one focusing on organizations) published in peer-reviewed publications ranging from conference proceedings to journal articles, as listed in Table 1. Paper A and paper B compare the advice utilization of AI-based services in response to RQ1. Paper A addresses high-stakes decisions by choosing financial robo-advisors as the scenario. In contrast, paper B addresses low-stakes decisions by choosing a guessing game as the scenario. Through these two research papers, this thesis examined RQ1 from two perspectives, for low- and high-stakes decisions. In response to RQ2, paper C determined the acceptance factors of older people of health chatbots to extend research that previously investigated the acceptance factors for younger people. Paper D compared the effectiveness of eleven measures to increase trust in a financial robo-advisor as an AI-based service and, therefore, responds to RQ3. Paper E addresses RQ4 by analyzing the impact on organizational learning of the implementation of AI-based agents and describes AI-based services within an organization.

Table 1. Publications of this Thesis.

Paper A	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 B	Mesbah, Neda; Tauchert, Christoph; Buxmann, Peter (2021): **Whose Advice Counts More – Man or Machine? An Experimental Investigation of AI-based Advice Utilization.** In: 54th Hawaii International Conference on System Sciences (HICSS-54), Online, VHB-Ranking: C.
Paper C	Mesbah, Neda; Pumplun, Luisa (2020): **"Hello, I'm here to help you" – Medical care where it is needed most: Seniors' acceptance of health chatbots.** In: European Conference on Information Systems (ECIS), Online, VHB-Ranking: B.
Paper D	Mesbah, Neda; Tauchert, Christoph; Olt, Christian; Buxmann, Peter (2019): **Promoting Trust in AI-based Expert Systems.** In: Americas Conference on Information Systems (AMCIS), Cancun, Mexico, VHB-Ranking: D.
Paper E	Sturm, Timo; Gerlach, Jin P.; Pumplun, Luisa; Mesbah, Neda; Peters, Felix; Tauchert, Christoph; Ning, Nan; Buxmann, Peter (2021): **Coordinating Human and Machine Learning for Effective Organizational Learning.** In: MIS Quarterly 45(3), p. 1581 - 1602, VHB-Ranking: A+.

The five publications included in this thesis use different research designs (see Table 2, column 2). Besides papers A, B, and D, which are based on quantitative research methods (some of them used an experimental design) to measure users' advice utilization or the methods by which their perception of trust increased, paper C is based on a qualitative research method using semi-structured interviews with older people. Paper E is based on a simulation to reveal the impact of AI-based agents on organizational learning. In addition, in papers A and B a qualitative preliminary study was conducted. Therefore, these two papers used the advantages of the exploratory nature of qualitative studies (Mingers 2001; Venkatesh et al. 2013) with an open-ended survey.

Table 2. Outline of Research Papers.

Chapter and Research Paper	Research Type and Methodology	Theoretical Background	Context
Chapter 3 Research Paper A: Acceptance and Utilization of AI-based Services for High-Stakes Decisions	Qualitative Pre-Study; Experimental Survey Study	Judge-Advisor System; Task-Technology Fit	Financial Robo Advisor
Chapter 4 Research Paper B: Acceptance and Utilization of AI-based Services for Low-Stakes Decisions	Qualitative Pre-Study; Experimental Survey Study	Judge-Advisor System; Task-Technology Fit	Guessing Game
Chapter 5 Research Paper C: User-Group-Specific Acceptance and Utilization of AI-based Services	Interviews	Unified Theory of Acceptance and Use of Technology (UTAUT2)	Health Chatbot
Chapter 6 Research Paper D: Promotion of AI-based Services	Survey Study	Adoption of Innovations; Judge-Advisor System; Technology Acceptance	Financial Robo Advisor
Chapter 7 Research Paper E: AI-based Services in Organizations	Simulation	Organizational Learning	AI Agents in an Organization

Besides the publications included in this cumulative dissertation (see Table 1), I co-authored the following peer-reviewed publications as a PhD student at the Technical University of Darmstadt, Germany:

- Wagner, Amina; Mesbah, Neda (2019): Too Confident to Care: Investigating Overconfidence in Privacy Decision Making. In: European Conference on Information Systems (ECIS), Stockholm, Sweden, VHB-Ranking: B.
- Olt, Christian; Mesbah, Neda (2019): Weary of Watching Out? – Cause and Effect of Security Fatigue. In: European Conference on Information Systems (ECIS), Stockholm, Sweden, VHB-Ranking: B.
- Tauchert, Christoph; Bender, Marco; Mesbah, Neda; Buxmann, Peter (2020): Towards an Integrative Approach for Automated Literature Reviews Using Machine Learning. In: 53rd

Hawaii International Conference on System Sciences (HICSS-53), Wailea, Maui, Hawaii, VHB-Ranking: C.

Chapters 3 to 7 contain all publications included in this thesis[2]. Chapter 2 defines the theoretical fundamentals of the thesis. The thesis concludes with a summary of the overarching key findings and the conclusion is in Chapter 8. Figure 1 shows the structure of this thesis.

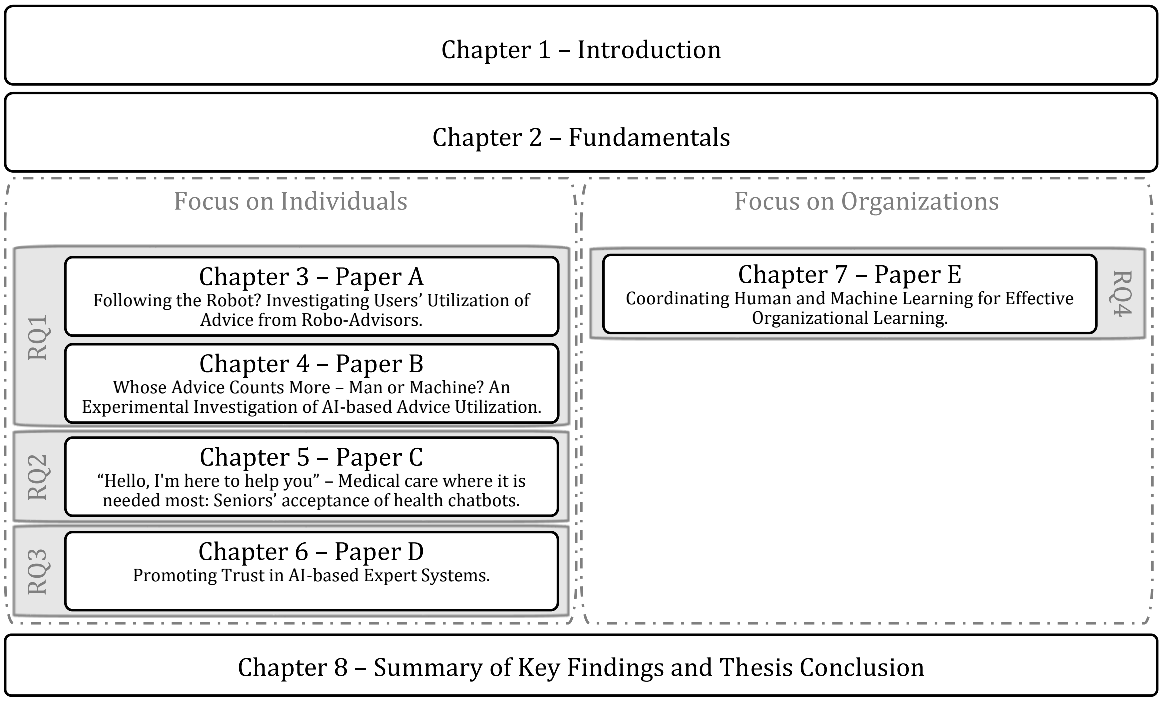

Figure 1. Structure of the Thesis.

[2] In order to ensure a consistent layout in this thesis, the papers have been slightly adapted from their original versions. They have also been written in the first person plural (i.e., "we"), as co-authors were involved in each publication.

2 Theoretical Background

This chapter explains the underlying theories that form the basis of and are used in the research of this cumulative dissertation. In particular, the concepts of artificial intelligence, machine learning, and the concept of services as well as AI-based services are introduced, along with models of technology acceptance. The chapter concludes with an explanation of organizational learning.

Artificial Intelligence

The term "AI" describes science and engineering that concerns learning and perception in machines that imitates human intelligence using underlying technologies such as machine learning, deep learning, and natural language processing (Elliot and Andrews 2017; McCarthy 2007). There is no common definition of AI-based systems in the literature. For example, the definition by Russell and Norvig (2016) is based on two dimensions. Their first dimension refers to thinking processes and the behavior of AI-based systems, and the second dimension describes how the success of such a system is measured, specifically whether success is measured against human performance or an ideal (rational) performance. This leads to four characteristics that define an AI-based system: thinking humanly, thinking rationally, acting humanly, or acting rationally (Russell and Norvig 2016). A characteristic of AI-based systems is that they do not follow any predefined static rules, in comparison to other technologies, but learn based on data (Burrell 2016).

One current approach that is widely used to implement data-based learning is machine learning (ML), which derives patterns from data using learning algorithms to create models (Brynjolfsson and Mitchell 2017; Jordan and Mitchell 2015; Mitchell 1997; Russell and Norvig 2016). The operationalization is possible in a variety of ways, for example through supervised, unsupervised, and reinforcement ML These operationalizations differ in that they use different types of learning algorithms to derive patterns in data. For example, supervised learning uses labeled data, and therefore it is known at which input which output is to be expected. This has the advantage that it is obvious what should be learned. In addition, it is possible to validate the accuracy of the model using data. In practice, however, labeled data are often not existent and can be very expensive and time-consuming to create (Masud et al. 2008). In contrast, unsupervised learning uses unlabeled data. The advantage is that a model can be created without labeled data by grouping similar data points. However, there is no direct influence on the data categorization and this can lead to the creation of an undesired grouping. Each form of ML creates a model based on data and learning algorithms (e.g., Mitchell 1997).

Services Based on Artificial Intelligence

Due to developments in technology and the increased use of ML techniques, systems based on AI can now be used for many tasks that were previously only performed by humans (e.g., Seeger et al. 2018). There are an increasing number of services that are AI-based. Services are (intangible) activities that aim to solve customer problems usually in customer interactions (Grönroos 2001).

A popular example of an AI-based service is financial robo-advisors (Jung, Dorner, Glaser, et al. 2018; Jung and Weinhardt 2018; Sironi 2016a), which are automated investment advisory services. These services aim to recommend target-oriented investment strategies. Customers are guided through a self-assessment process and possible portfolio compositions to estimate stock performance (Jung, Dorner, Glaser, et al. 2018; Sironi 2016a; Tertilt and Scholz 2017). Robo-advisors have benefits for providers and users. Providers have a scalable advisory service that reduces their investment costs. Users can also reduce their investment costs, and they can monitor their portfolios in real-time because the robo-advisors are always available and process substantially more information than human advisors due to humans' cognitive restrictions (Simon 1972; Tertilt and Scholz 2017). Obviously, there are also some deficiencies. Providers cannot build close relationships with customers due to the lack of interactions, and users cannot discuss investment strategies as they would with a human advisor (Samek et al. 2017; Wanner et al. 2020). Specifically, the robo-advisor models that are used for recommendations are often not interpretable by the provider (i.e., the developer) or the user (Lipton 2016), but humans can conduct a dialogue so that the recommended investment strategy can be questioned and explained.

Another popular example of an AI-based service is chatbots. Chatbots are text-based conversation systems that interact with users using natural language and emulating human-to-human conversations (Seeger et al. 2018; Shawar et al. 2005). Depending on the scope of the application, they can be divided into general-purpose or domain-specific bots (Gnewuch et al. 2017). Domain-specific bots can be found in a wide variety of areas, such as museums (e.g., Kopp et al. 2005), e-commerce (e.g., Qiu and Benbasat 2009), or the healthcare sector (e.g., Minutolo et al. 2017). Depending on the application's field of use, there are benefits and drawbacks for both the provider and the user of a chatbot. However, it can generally be said that from the provider's perspective, chatbots offer the ability to automate basic, repeatable, standardized customer-service interactions, eliminating the need for human employees to handle these interactions, which can save significant labor costs (Fernandes 2017). One of the benefits for users of chatbots is that these chatbots are always accessible so users' questions or problems can be solved at any time of day conveniently and quickly (Brandtzaeg and Følstad 2017). On the other hand, it is difficult for providers to build close customer relationships, as negative experiences can also quickly arise when using chatbots (Følstad and Brandtzaeg 2020). Specifically, misunderstandings can arise that are more difficult to resolve than in human-to-human interactions, and this is a detriment from the user's perspective (Følstad and Brandtzaeg 2020).

There are many other examples of such AI-based services. However, users who have been used to interacting with and trusting humans in the past may not accept and use such services. Different results regarding the acceptance and use of AI-based services can be found in the literature (Jussupow et al. 2020). Some of them show that the services are accepted and used, and even preferred, in contrast to human services (e.g., Logg et al. 2019) and others have found contradictory results (e.g., Castelo et al. 2019; Lee 2018).

Individual Acceptance and Utilization of Technology

Considering these opposing findings, it becomes apparent that further and more in-depth research is needed. Regarding the acceptance and use of IT systems in general, information system (IS) research has already developed several theories and models, which can be used for further research.

One of the fundamental models in IS research is the technology acceptance model (TAM) (Davis 1989). The model assumes the actual system use is influenced by the behavioral intention to use the system and the behavioral intention is influenced by the attitude toward the system. Moreover, the attitude toward the system is affected by two factors: the perceived usefulness (i.e., to what extent the user believes the system is useful in achieving the goal of use) and the perceived ease of use (i.e., how easy the user perceives the system to be to use). Finally, these factors are also influenced by external variables like social influence (see Figure 2). This model has subsequently been developed several times and empirically verified in various studies (Lu and Yang 2014).

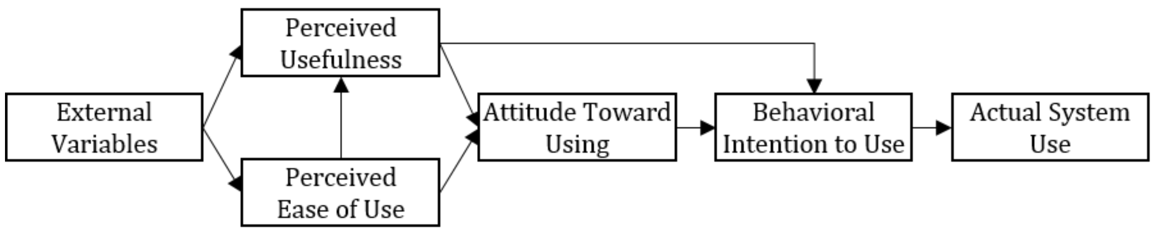

Figure 2. Technology Acceptance Model.
(Based on Davis 1989)

Another very well-known theory phenomenon is the task-technology fit (TTF), which assumes a technology and its corresponding technological characteristics must fit to solve a specific task and its characteristics (e.g., difficulty, significance, routineness) to be accepted by users (Goodhue 1995; Goodhue and Thompson 1995). Such a fit is expected to serve as a predictor of the utilization of an information system, as individuals are more likely to use a technology that they perceive to be appropriate to solve a task (Goodhue and Thompson 1995). Figure 3 illustrates the model. The TTF is also verified in various studies and various contexts (e.g., Goodhue et al. 2000; Klopping and Mckinney 2004; Robles-Flores and Roussinov 2012; Zigurs et al. 1999; Zigurs and Buckland 1998).

Figure 3. Task-Technology Fit.
(Based on Goodhue and Thompson 1995)

The Unified Theory of Acceptance and Use of Technology is another recently developed and widely used model in the consumer context (UTAUT) (Venkatesh et al. 2003). Similar to TAM, UTAUT assumes that behavior is influenced by usage intention and introduces four influencing factors: performance expectancy (i.e., “the degree to which using a technology will provide benefits to consumers in performing certain activities” (Venkatesh et al. 2012, p. 159)), effort expectancy (“the degree of ease associated with consumers’ use of technology” (Venkatesh et al. 2012, p. 159)), social influence (“the extent to which consumers perceive that important others (e.g., family and friends) believe they should use a particular technology” (Venkatesh et al. 2012, p. 159)), and facilitating conditions (“consumers’ perceptions of the resources and support available to perform a behavior” (Venkatesh et al. 2012, p. 159)) (see Figure 4). This model is also extended as UTAUT2 and verified in various studies (e.g., Laumer et al. 2019; Nguyen et al. 2015; Sell et al. 2017).

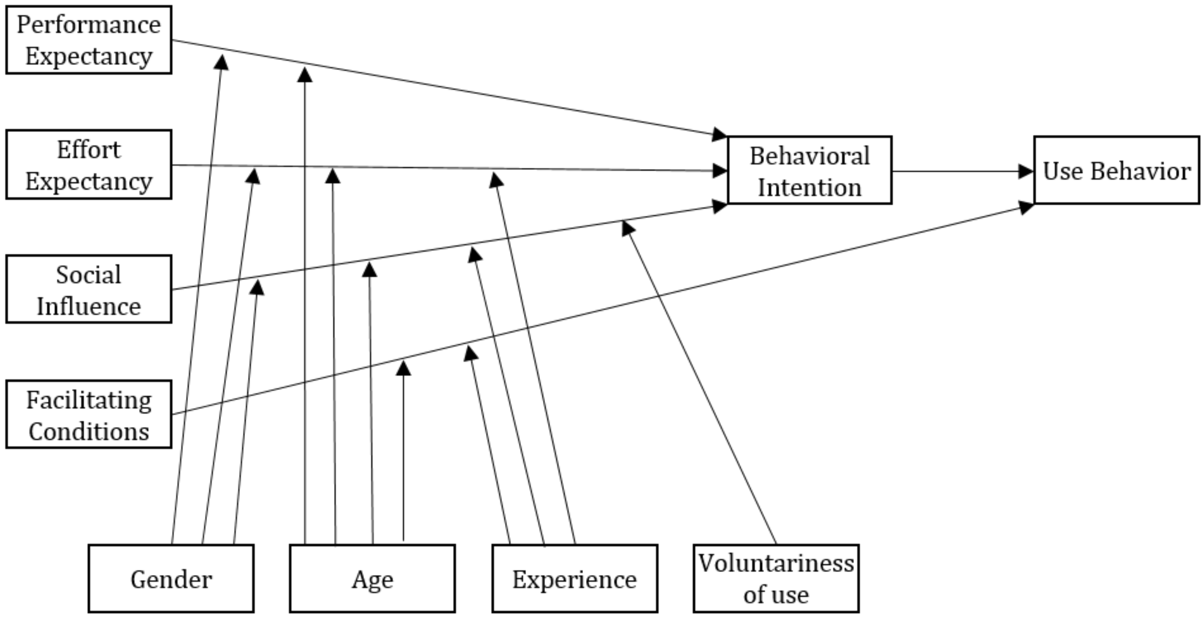

Figure 4. Unified Theory of Acceptance and Use of Technology.
(Based on Venkatesh et al. 2003)

In addition to these three basic models, there are many extensions and other models concerning the acceptance and utilization of IT systems in general. Although all of the models have been verified in various studies, the setting of AI-based services is very recent and has new usage behavior, especially because AI-based services replace services that could previously only be performed by humans. Therefore, these models need to be verified again to determine whether they fit in the AI-based services context or need to be extended.

Organizational Learning

Once AI-based services are accepted and used by members within an organization, these services represent another source of knowledge, which is why the use of AI-based services should also be evaluated from the perspective of organizational learning.

Organizations learn by creating routines from inferences derived from history, by which they are guided (Levitt and March 1988). These routines form the organizational memory and are independent of the knowledge levels of individual members of an organization (e.g., Feldman and Pentland 2003; Levitt and March 1988). Nevertheless, organizational knowledge is based on the beliefs of the individuals within an organization. The more diverse the beliefs, the more effective organizational learning is from a long-term perspective (March 1991). In addition, higher levels of organizational knowledge indicate a higher probability of competitive advantage, which makes organizational learning important (Hatch and Dyer 2004; Levinthal and March 1993; March 1991).

Thus, it is important for an organization to maintain the diversity of individual beliefs for as long as possible. High levels of diversity of beliefs allow organizations to experiment with alternatives before committing to a particular solution. The hurdle is that is also important for organizations to disseminate correct beliefs and prevent the spread of false beliefs. However, organizations should avoid prematurely disseminating correct beliefs, as this could lead to being trapped in inferior routines (Denrell and March 2001; Levinthal and March 1993; March 1991). An appropriate balance between diversity and the use of existing correct beliefs is crucial for an organization's survival. This means that managing diversity of beliefs through the coordination of individual learning within the organization is essential to effective organizational learning (e.g., Lavie et al. 2010; March 1991).

One way to manage the diversity of beliefs is to influence the pace of the homogenization of individuals' beliefs. To accomplish this, organizations must balance exploration with exploitation (e.g., Lavie et al. 2010; March 1991). Exploration preserves the diversity of beliefs by encouraging the search for new ideas that may lead to uncertain outcomes. Exploitation accelerates the homogenization of beliefs by focusing on existing knowledge and incremental improvements (e.g., Lavie et al. 2010; March 1991; Raisch et al. 2009).

3 Research Paper A: Acceptance and Utilization of AI-based Services for High-Stakes Decisions

Title: Following the Robot?
Investigating Users' Utilization of Advice from Robo Advisors

Authors: Tauchert, Christoph; Mesbah, Neda

Published in: International Conference on Information Systems (ICIS), Munich, Germany, 2019

Abstract

Companies are gradually creating new services such as robo-advisors (RoA). However, little is known if users actually follow RoA advice, how much the fit of RoA to task requirements influences the utilization, how users perceive RoA 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 RoA. While the perceived advisor's expertise is the most influential factor on task-advisor fit for RoA and human advisors, integrity is a significant factor only for human advisors. However, for RoA the user's perception of the ability to make decisions efficiently is significant. In our study, users followed RoA 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.

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. Russell and Norvig (2016) 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 (Russell and Norvig 2016). 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 2016a). 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 2016a; Tertilt and Scholz 2017). If robo-advisors were accepted by users[3], 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

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

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 1995). 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 (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.

Theoretical Background

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.

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 1995).

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 1995). Figure 5 depicts the general idea of the model.

Figure 5. 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.

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 1995), 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 6 shows the final research model, which builds the foundation of our study.

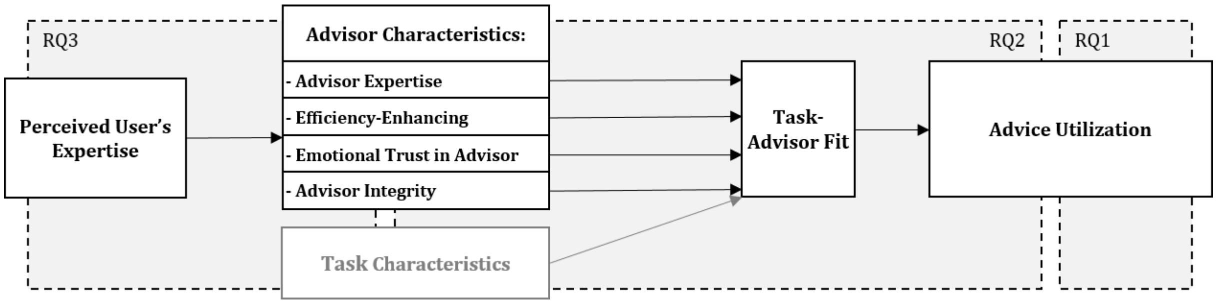

Figure 6. Research Model.

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).

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.

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 2004a). 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 1. 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).

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).

Table 3. Item Loadings.

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 Efficient-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

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 3 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 4.

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 4, 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.

Table 4. Cronbach's α (Cr. α), Composite Reliability (CR), Average Variance Extracted (AVE) and Construct Correlations.

(First Row: Robo-Advisor; Second Row: Human Advisor)

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

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 5.

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 andH6b. All other hypotheses had to be rejected. These findings of our multi-group analysis are visualized in Figure 7.

Table 5. Results of Structural Model Testing and Effect Sizes.

*(*** p<0.001; ** p<0.01; * p<0.05; RoA=Robo-Advisor, HU=Human Advisor)*

Constructs	**Path Coefficients and p-Values**				**Multi-Group Testing**		**f^2 values**	
	RoA	**p**	**HU**	**p**	**Diff.**	**p**	**RoA**	**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

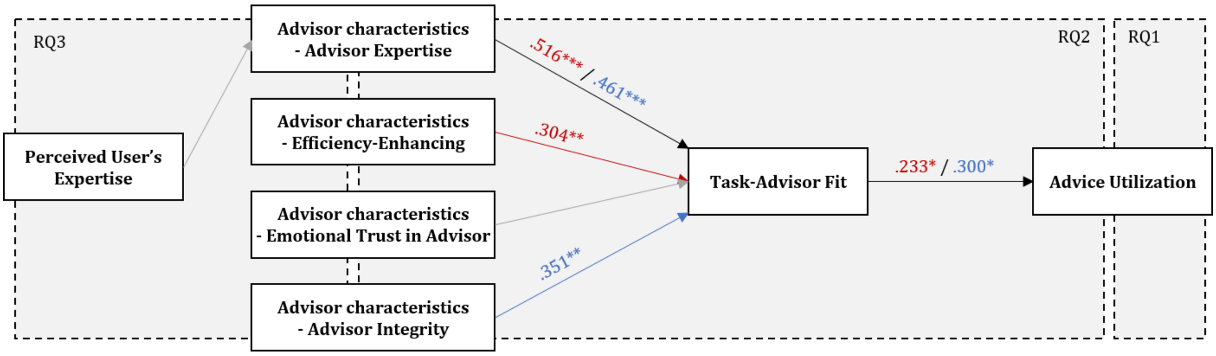

Figure 7. Summary of Structural Model Testing.
(Robo-Advisor in Red and Human Advisors in Blue).

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.

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.

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.

Acknowledgments

This research project was funded by the Hessian state ministry "Hessisches Ministerium des Innern und für Sport" in Germany.

4 Research Paper B: Acceptance and Utilization of AI-based Services for Low-Stakes Decisions

Title: Whose Advice Counts More –
Man or Machine? An Experimental Investigation of AI-based Advice Utilization

Authors: Mesbah, Neda; Tauchert, Christoph; Buxmann, Peter

Published in: 54^{th} Hawaii International Conference on System Sciences (HICSS-54),
Online, 2021

Abstract

Due to advances in Artificial Intelligence (AI), it is possible to provide advisory services without human advisors. Derived from judge-advisor system literature, we examined differences in the advice utilization depending on whether it is given by an AI-based or human advisor and the similarity of the advice and their own estimation. Drawing on task-technology fit we investigated the relationship between task, advisor and advice utilization. In study A we measured the actual advice utilization within a guessing game and in study B we measured the perceived task-advisor fit for this game. The findings show that compared to human advisors, judges utilize advices of AI-based advisors more when the advice is similar to their own estimation. When the advice is very different to their estimation, the advices are used equally. Concluding, we investigated AI-based advice utilization and presented insights for professionals providing AI-based advisory services.

Keywords: Artificial Intelligence-Based Assistants, Artificial Intelligence, Advice Taking, Judge-Advisor System, Distance Effects, Task-Technology Fit.

Introduction

Decisions are part of our everyday lives. How many decisions do you think you make per day? An average adult makes about 226 decisions every day – just about food (Wansink and Sobal 2007), and probably tens of thousands in general. Many decisions are not made alone, but are discussed with other people like parents, friends or experts. In particular, experts provide important decision-making assistance in the event of uncertainties due to a lack of personal knowledge or experience (Sniezek and Van Swol 2001). Technological development has enabled not only human experts to support us in decision-making based on their knowledge and experience, but also machines based on artificial intelligence (AI).

A common definition describes AI as "science and engineering of making intelligent machines, especially intelligent computer programs" through a simulation of human intelligence by underlying technologies like machine learning, deep learning and natural language processing (Elliot and Andrews 2017; McCarthy 2007, p. 2). AI differs significantly from other technologies, since these AI-based systems have the ability to learn and not just follow static rules (Burrell 2016).

AI-based advisors are often called robo-advisors. They are, compared to humans, only machines that simulate the learning abilities of humans (McCarthy 2007) but not with the same interaction possibilities as with a human advisor. During the interaction with a robo-advisor, the decision-maker gets a target-oriented advice based on a previous self-assessment process (Jung, Dorner, Glaser, et al. 2018; Sironi 2016b). Often the models used to generate an advice are not interpretable by neither the user nor the developer (Lipton 2016). In contrast, human beings can be engaged in dialogue and an advice can be questioned and explained. However, there are some advantages of using AI-based experts, such as that they are able to process much more information than humans who are cognitively restricted (Simon 1972) or that they are always available. Hence, the question arises whether the differences between AI-based and human advisors also lead to different utilization of their advice.

For example, in a study by Tauchert and Mesbah (2019) participants preferred the advice of a financial robo-advisor over that of a human advisor. In the literature we can find different findings about the utilization of AI-based experts (Jussupow et al. 2020). Some studies show that AI-based systems are preferred in contrast to human experts (e.g., Logg et al. 2019; Tauchert and Mesbah 2019) and other find contradictory results (e.g., Castelo et al. 2019; Lee 2018).

However, it is not clear whether this different utilization of an advice is also present in other contexts and if the preference is predictable. Moreover, this different utilization could be affected by characteristics of the advice. As soon as an advice is given, the decision-maker perceives compulsorily advice characteristics and connects them to the advisor. Thus, the literature shows that particularly the similarity of the advice given to one's own estimation has a great influence on the degree of advice utilization (Minson et al. 2011; Yaniv 2004a). This leads us to the following research questions:

RQ1: Do people utilize advice differently depending on whether it is given by human or artificial intelligence and is the different utilization predictable?

RQ2: Is the different advice utilization of human and artificial intelligence advisors depending on the distance of the advice to their own estimation?

In Information System (IS) literature, the task-technology tit (TTF) is used to determine how well a technology is suited to assist a person performing a task (Goodhue and Thompson 1995). By following the approach of Tauchert and Mesbah (2019) and adopting this model in the judge-advisor context, it would be possible to combine all the factors so far considered in the judge-advisor system (JAS) literature and to create a holistic view. This model could be used to predict, whether a human or AI-based advisor is followed more depending on the task.

By answering these research questions we also follow the call of Rzepka and Berger (2018) for investigations about the user's[4] utilization behavior of AI-based systems. Specifically, they have highlighted that there is still little research on AI-based advice. Therefore, we conducted an online experimental survey.

Advice Utilization

People use advice for three main reasons: improvement of their judgement, sharing of responsibility and simply refusal to completely reject received advice (Harvey and Fischer 1997). One paradigm that is used in behavioral psychology to investigate advice-taking behavior is the judge-advisor system (Sniezek and Buckley 1995). It is a structured group in which one group member, the judge or decision-maker, seeks out advice from one or more advisors (which can be an expert or not) and can aggregate the advice with their own judgment (Van Swol 2011). The utilization of advice or also called weight of advice is defined as the relative adjustment of a decision-maker from his initial advice towards the advice they receive from an advisor (Bonaccio and Dalal 2006).

However, we have to consider different factors – such as trust, advisor's competence, distance of advice, expertise of judge, task difficulty – which influence advice-taking behavior (Gino and Moore 2007; Schultze et al. 2015; Sniezek and Buckley 1995; Van Swol and Sniezek 2005; Yaniv 2004a). A summary of JAS studies can be found in the paper of Bonaccio and Dalal (2006). The factors can be categorized in four clusters: characteristics of advisor, characteristics of judge,

[4] In our context users are decision-makers and users of an advisory services.

characteristics of task and characteristics of advice. Since we want to measure how AI-based experts are perceived in comparison to human experts, we primarily focus on advisor and advice characteristics in this manuscript.

There are several different advisor characteristics in the JAS literature discussed. Some of these factors such as similarity to decision-maker (Gino et al. 2009) and age (Feng and MacGeorge 2006) are not transferable to an AI-based expert. Therefore, we will only consider the factors that can be perceived in both, a human and an AI-based system. One of the most discussed factors is trust. 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). There is not one definition but many different, but by now most researchers agree that it is a multidimensional concept (Komiak and Benbasat 2006; Rousseau et al. 1998). Mayer et al. (1995) categorized trust in competence, integrity and benevolence. By following this definition, several studies investigate the impact of advisor's expertise on advice utilization. Advisor's expertise is the perceived ability of the advisor to give a good advice in a specific domain (Mayer et al. 1995). The more competent an advisor is perceived, the more willing a judge is to adjust his estimation (e.g., Kim et al. 2017; Schultze et al. 2015). Integrity is defined as the advisor's honesty and promise keeping (McKnight et al. 2002). The higher the perceived integrity of the advisor is, the more likely it is that the advice will be used (Van Swol 2011). The same applies to the perceived benevolence of an advisor (White 2005), which describes how much an advisor cares about the judge and acts in his interest (McKnight et al. 2002).

So far, we have only considered factors that have been examined with human advisors. However, the JAS literature also examines some factors that are particularly relevant to the use of non-human advisors, especially for recommender systems. One of the most discussed factors in this area is the ability to provide explanation. Several studies show that the justification of a recommendation is effective in changing users attitude towards the usage of an advice (Wang and Benbasat 2007; Ye and Johnson 1995; Zanker 2012).

Studies which focus on advice characteristics have also shown that advice utilization differs based on the gap between the decision-makers' and advisors' opinion, called distance of advice. When the advice is similar to the decision-maker's own estimation, the distance is close, whereas when the advisor gives a completely different advice compared to the decision-maker's estimation the advice distance is far. While at first there was evidence for a monotone negative relation of advice distance and advice utilization (i.e., advice is weighted more when it's close to the decision-maker's opinion and less when it is far from it) (Minson et al. 2011; Yaniv 2004a), a more recent study shows that it might be a more complex relation. Schultze et al. (2015) find a curvilinear pattern, where advice is weighted less when advice distance is too low as well as too high. This can be explained by the effect of social validation, meaning that a perceived similar opinion increases the decision-maker's confidence in his beliefs leading to non-adaption of the already similar advice.

Summarized, the JAS literature identifies some factors that can influence advice utilization. In order to investigate whether advice from a human or AI-based advisor is perceived differently, we want to adapt the task-technology fit, which will be described in the next section, in the judge-advisor context.

Task-Technology Fit

The TTF was initially introduced in IS literature to investigate the relationship between information systems and an individual's performance. Goodhue and Thompson (1995) extend the TTF to the Technology-to-Performance Chain and they showed that TTF has a direct impact on the utilization of an IT system as well as on individual performance. TTF is defined as "the degree to which a technology assists an individual in performing his or her portfolio of tasks" (Goodhue and Thompson 1995). For instance, in the case of a high TTF, the capabilities of the technology match the requirements of the task very well. Technologies are all kinds of tools from computer systems to support services that can help an individual to carry out a task. By employing such a technology during the task solving process, this technology will be utilized. If a system will be used or not depends on individual beliefs about the consequences of usage. The TTF reflects these beliefs, i.e., the TTF reflects if a user believes the technology has any relative advantages. In conclusion, this linkage implies the impact of TFF on utilization. Several studies have validated the TTF model in different contexts such as question-answering system or group support systems (Robles-Flores and Roussinov 2012; Zigurs and Buckland 1998). Next, we adopt the TTF model in the JAS context and derive our hypotheses.

Research Model

Until now, the JAS was mainly utilized to investigate the interaction between human decision-makers and human advisors (Gino and Moore 2007; Sniezek and Buckley 1995; Sniezek and Van Swol 2001). However, there is one study that investigates differences in the utilization of advice when using a statistical model compared to human advice (Önkal et al. 2009). They showed that decision-makers discount statistical advices more than human advices. The participants weigh an advice differently just because they perceive a different source although the advice is presented in the exact same way. However, due to the increasing amount of data and computer power, AI algorithms are used nowadays in a constantly growing manner (Anthes 2017). Therefore, another study investigates differences in the utilization of advice of human compared to financial robo-advisors (Tauchert and Mesbah 2019). The participants utilized the advice of a financial robo-advisory more than a human advisor even though the advice is presented in the exact same way. Beside the JAS literature we find different findings in other research streams about the preference of AI-based advices (Jussupow et al. 2020). Some of the studies show an algorithm aversion, i.e. a preference of human advisors (e.g., Castelo et al. 2019; Lee 2018) while other studies show a preference of AI-based advisors (e.g., Logg et al. 2019; Tauchert and Mesbah 2019). It seems that the preference is depending on the task and advisor characteristics. Therefore, we hypothesized:

H1: AI-based expert advices will be differently utilized compared to human expert advices.

As described above, advice can be characterized by its distance to the initial estimation of the decision-maker. Depending on its distance an advice will be weighted differently. Schultze et al. (2015) have shown that whenever the advice of an advisor is far away, usually that leads to a change in our own estimation. Therefore, we assume if an advice is far enough away, the difference between the characteristics of an AI-based and human expert will not be large enough to suppress the desire to adjust his estimation. This leads to the following hypothesis:

H2: The preference for AI-based or human advisor will decrease with increasing distance of advice.

By transferring the TTF to the JAS context, we want to measure the degree to which an advisor assists a decision-maker when performing a task, called task-advisor fit (TAF). Due to the different characteristics of AI-based and human advisors the perceived fit to a task should differ. As described above, TTF is a predictor of the utilization of IT systems (Goodhue and Thompson 1995). Therefore, TAF should be a predictor of the utilization of an advice, so we hypothesized:

H3: Perceived task-advisor fit reflects the advisor preference and advice utilization.

The TTF model shows that the technology characteristics have an impact on the perceived fit (Goodhue and Thompson 1995). Equivalently, we propose above identified advisor characteristics would contribute to the judges' perception of TAF. As described the expertise of an advisor affects the judges willingness to follow the advice (e.g., Kim et al. 2017; Schultze et al. 2015). It seems that the higher advisor's expertise is perceived, the more the advisor to the task fits. The same applies to the rest of above identified advisor characteristics integrity, benevolence and providing explanations. To ensure that we have covered all relevant advisor characteristics through the literature, we conducted a pre-test with 67 participants. We asked them to list characteristics which they associate with a human advisor, which they associate with an AI-based advisor and which differences they perceive. The result confirmed most of the literature advisor characteristics, such as competence and providing explanations. Based on the pre-test we added the efficiency-enhancing characteristic, that describes the extent to which an advisor enables efficient decision-making. Accordingly, we hypothesized:

H4a-e: The advisor's characteristics expertise, efficiency-enhancing, integrity, benevolence, providing explanations positively affect the TAF.

We visualize our research model in Figure 8. After we have derived the research model, the next chapter presents the research method we used to test our model.

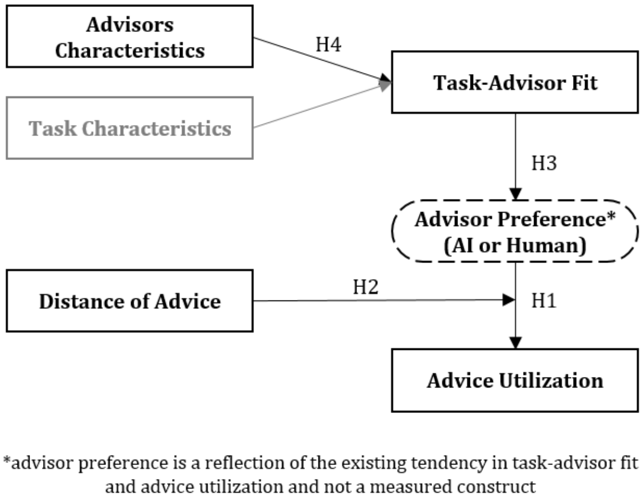

Figure 8. Research Model.

Research Method

To investigate if there are any differences between utilization of advice from AI-based experts compared to human experts, we set up two online survey studies. In study A we conducted an online experimental survey following the call of Rzepka and Berger (2018) to study user's actual advice utilization during the interaction with AI-based systems and not only the self-reported perception. During this experiment, we played a guessing game with the participants (see Guessing Game Description). The participants of the study had the chance to win up to 5€ during the game in order to evoke their actual behavior (Camerer and Hogarth 1999). In study B we conducted a scenario-based online survey to examine the TAF of the same guessing game and to identify the key advisor characteristics responsible for the fit. We chose to conduct two different online studies so that the actual implementation of the experts does not affect the perceived TAF and vice versa. With two different studies we can actually determine whether the TAF is a proxy for actual behavior. It is worth noting that in study B we just described the guessing game whereas in study A we actually played the game. Both studies have a 2x1 between-subjects design, i.e. we randomly divided the participants of the surveys into two groups (AI-based vs. human advisor). Both groups in both studies got a description of our guessing game following the approach of Gino and Moore (2007).

In order to ensure a high degree of representativeness of the population in terms of age, gender and occupation among internet users, the two surveys were conducted with the help of a market research company (for justification see Lowry et al. (2016)). At the beginning of each study, participants were made aware that the survey was anonymous and that there were no right or wrong answers besides the answers during the guessing game, to counteract the common method bias (Podsakoff et al. 2003). Afterwards, they were introduced to the game.

Guessing Game Description

We have described the game as follows: "You will hear a tic tac box being shaken. Your task is to estimate how many tic tacs are in the box.". All participants were told that an expert would provide his own estimation after they made an initial estimation. After they have received this information from the expert, the participants could adjust their estimation if desired. We told the first group that the expert was an AI-based system and the second group was a human. The AI-based expert was described as an application based on artificial intelligence and explicitly trained to estimate the number of tic tacs in a box, which performs well. Instead, the human expert was described as an expert who has perfect pitch and has explicitly been trained to estimate the number of tic tacs, who performs well. Beyond that, the experts were identical and were only presented in this way. All participants were additionally informed, that the more precisely they estimate the true value, the more profit they get. At no time during the experiment, they received an information about the true number of tic tacs contained in the box.

Each participant played eight rounds in our game. How well they performed and how much they won was only revealed to them at the end of the study. Further provided information consisted in the maximum number of 37 tic tacs that would fit in a tic tac box to ensure that all participants had the same knowledge base. All in all, one round of the game consisted of the following steps:

1. Participants listened to the audio file. The audio file contained a tic tac box being shaken.
2. Initial estimation: The participants estimated the number of tic tacs in the box.
3. Participants received additional information from an AI-based expert or a human expert.
4. Final estimation: Participants were able to adjust their estimation if desired.
5. The next round began.

Four different amounts of tic tacs were used in the eight rounds (each with 7, 11, 16 and 29 tic tacs), i.e., every participant heard every amount of tic tacs twice. However, to ensure that the participants did not recognize audio sequences were being repeated, we recorded new audio files for each round.

In order to check whether it makes a difference how close the expert's advice is to the initial estimation of the judge, a manipulation was carried out. For each amount of tic tacs (7, 11, 16 and 29) the advisor gave close (CA) and far advice (FA) comparing to the initial estimation of the participant. Close advices had a maximum difference of three tic tacs compared to the initial estimation whereas far advices had a difference of at least seven tic tacs. The participants played 8 rounds with the following sequence of: (1) 7 tic tacs with CA, (2) 16 tic tacs with FA, (3) 11 tic tacs with CA, (4) 29 tic tacs with FA, (5) 11 tic tacs with FA, (6) 29 tic tacs with CA, (7) 7 tic tacs with FA and (8) 16 tic tacs with CA.

The amount of rounds, the amount of tic tacs, the distance between the expert's advice and one's own initial estimation, as well as the order of the rounds were pre-tested in a laboratory experiment (n=27). For the amount of tic tacs and the distribution over the rounds, we ensured that they were evenly distributed and that the participants did not assume that the advice had been manipulated. The distance between advice and initial estimation was developed on the basis of the results from Schultze et al. (2015), so that we ensured that participants had perceived small or large deviations from their estimates as such. Finally, we have taken care to select the number of rounds so that the participants could still process all the information provided to them.

We chose to guess tic tacs by listening to an audio file for five main reasons: (1) The game is very intuitive and easy to understand. (2) The probability that participants are confident in their own estimates is low, as their experience might be low. Therefore, the advice should be helpful. (3) It is easy to imagine that experts can estimate the number of tic tacs well through sufficient training. (4) It is easy to imagine that people with absolute pitch have advantages in being able to recognize and distinguish certain tones whereby they can perform this task well. (5) Finally, it is conceivable that an AI is able to recognize patterns with the help of machine learning and thus fulfil this task well.

After the participants were introduced to this guessing game, we presented them the items of our main constructs in study B and in study A they start to play.

Items Study A

To measure the degree of advice utilization we used the "weight of advice" (WOA), which has been used in serval 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 weight of advice is a measure that determines to what extent participants consider (weight) an advice in their estimation (Yaniv 2004a). If a participant completely ignores the advice and does not adjust his/her estimate, then the WOA is 0. On the other hand, if a participant completely adjusts his/her estimate to the advice, then WOA equals 1. A value for WOA between 0 and 1 means that a participant has partially adjusted his/her estimate to the advice, whereby a value of 0.5 means that a participant has formed the mean between his/her initial estimate and the advice.

Items Study B

Our main constructs in study B consist of TAF as well as advisor characteristics which we surveyed directly after the guessing game description whereas in study A we measured the actual advice utilization.

All our items were measured on a 7-point Likert scale ranging from ‘strongly disagree’ to ‘strongly agree’. To measure TAF we adopted the three items scale of Moore and Benbasat (1991) with statements like “The expert’s advisory service is compatible with all aspects of this task.”. For the evaluation of trust in integrity of advisors we applied the established scales of Komiak and Benbasat (2006) using three items with statements like “The expert is honest.”. Similarly, we measured the trust in advisor’s expertise based on a four item scale of McKnight et al. (2002) and trust in benevolence of advisor based on a four item scale of Kettinger and Lee (1997). We asked how much the participant agrees with statements like "The expert is competent and effective in estimating the amount of tic tacs." for expertise and for benevolence with statements like “The expert has your best interests at heart.”. To evaluate efficiency-enhancing we adopted the single item scale of Chan et al. (1997): “The expert increases the efficiency of my decision making.”. For the measurement of the ability to provide a justification of an advice we used the item “The advice I get from the expert is easy to comprehend.” from Zimmer et al. (2007). All of our construct measurements can be found in Appendix 3. We also measured tendency towards fantasizing as marker variable to counteract common method bias (Podsakoff et al. 2003) based on three item scale of Darrat et al. (2016).

Results of Study A

A total of 252 participants took part in study A. In order to guarantee the quality of the study results, we included an attention check to our survey and identified participants who gave the same answer across all constructs, so-called straight-liners (Maniaci and Rogge 2014; Meade and Craig 2012). After the exclusion of all straight-liners as well as all participants who failed in the attention check, 198 participants were left for further analysis. 47% of the study participants were female. On average, they were 37.81 years old (in a range of 18 to 69 years). Most participants were employees (59.6%), followed by students (13.6%). This corresponds almost to the European internet users’ distribution by age, gender and employment status (Eurostat 2018). 103 participants were assigned to the human expert group and 95 to the AI-based expert group. To compare the two groups with each other, we first ensured that the groups are equally distributed in their initial estimations. The average distance between the initial estimations and the advice does not significantly differ (approx. mean of 6 tic tacs).

Each participant of study 1 took part in 8 rounds of our game. This results in 1584 valid data points for the WOA measure. We followed the common procedure from the established literature (Gino and Moore 2007; Yaniv 2004b, 2004a) and replaced all values for WOA greater than 1 with 1. This is the case where the final evaluation is not within the range of advice and one's initial estimation. We have applied this to 1,89% (15 out of 792) of cases in the close advice condition and 2,15% (17 out of 792) of cases in the far advice condition. For each condition as well as for the total sample we calculated the mean of the WOA values and used them for further analysis.

To evaluate if there are any differences in the advice usage depending on whether the advice provider was an AI or a human being we ran an independent t-test and the results are reported in Table 6. There is a significant difference of advice utilization between the two groups. Participants adjusted their assessment more when the advice came from the AI-based expert rather than from a human expert, supporting H1. We also measured the perceived advisor expertise and investigated whether it is perceived differently in both groups. As the results of the t-test show (see Table 6), the AI-based advisor is perceived significantly more competent.

Table 6. Results of T-Tests for WOA and Advisor's Expertise Constructs.

Construct	**AI-based Advisor**		**Human Advisor**		**t-Test**			**Effect size**
	Mean	**SD**	**Mean**	**SD**	**t-value**	**df**	**p-value**	g_{Hedges}
WOA	.34	.235	.26	.173	2.688	171.607	.008(H1)	-.387
WOA_close	.28	.262	.17	.191	3.435	171.288	.001(H2)	-.495
WOA_far	.40	.263	.36	.238	1.297	196	.196(H2)	-.185
Advisor Expertise	4.72	1.126	3.99	1.190	4.170	176	.000	-.626

H2 postulates that by increasing distance of advice the impact of the preference of advisor decreases. By running independent t-tests we also check whether there were differences between the two groups with regard to close and far advice. Results of t-tests as well as the effect sizes are presented in Table 6. The analysis shows that there is no significant difference between the groups using far advice. However, participants who have received close advice from an AI-based expert utilized it significantly more than participants who have received close advice from a human expert. Consequently, H2 is supported.

Results of Study B

In study B a total of 265 internet users participated. To achieve a high quality of our study results, we implemented an attention as well as a manipulation check (Meade and Craig 2012). We excluded all participants who failed at least one check, who were too quick in answering the questionnaire as well as all participants who had never heard the term artificial intelligence or can't imagine what it means. After the exclusion 149 participants remained, 45% of whom were female. The age of the participants ranged from 18 to 68 years (mean age of 37.76 years) and most of them work as employees (57.7%), followed by students (11.4%). Our sample is again similarly distributed to the European internet users (Eurostat 2018). The sample size of the AI-based expert group is 89. We ensured that the groups are equally distributed in terms of age and gender.

H3 postulated that the TAF reflects the advisor preference and advice utilization. It is tested by running an independent t-test. The TAF of an AI-based advisor ($M = 4.58$, $SD = 1.551$) is statistically significantly higher than that of a human advisor ($M = 4.19$, $SD = 1.183$), $t(144.704) = 1.746$, $p = .042$, $d = .283$. Since both TAF and WOA show that AI-based advisors are preferred for this guessing game, H3 is supported.

To test H4a-e, we analyzed the impact of advisor characteristics on TAF. A well-established method for the analysis of such models are structural equation models as implemented in SmartPLS (Qureshi and Compeau 2009; Ringle et al. 2015). This suits well for theories in their early stages like ours (Fornell and Bookstein 1982).

To asses our measurement model we examined convergent and discriminant validity of the research model (Hair et al. 2013). Convergent validity ensures that items of the same construct are statistically similar. To confirm convergent validity, we evaluated item loadings, Cronbach's α and composite reliability (CR) and the average variance extracted (AVE) by the constructs (Xu et al. 2012). The item loadings were reported in Table 7. All items have higher loadings than 0.7 as recommended by Hair et al. (2013) so that our items are of sufficient reliability.

Table 7. Item Loadings.

Item	Item Loading	Item	Item Loading
TAF1	.884	INT1	.954
TAF2	.937	INT2	.962
TAF3	.936	INT3	.949
COM1	.955	SCOM1	.947
COM2	.972	SCOM2	.943
COM3	.964	SCOM3	.955
COM4	.942	SCOM4	.943
EFF	1.000	EXPL	1.000

As can be seen in Table 8, for all constructs Cronbach's α and composite reliability reach the threshold of 0.7 and AVE of 0.5 (Hair et al. 2011). The only exceptions are the Cronbach's α and AVE of the construct "Utilization of close advice", but due to the explorative nature of this study we consider these values acceptable (Codish and Ravid 2014). Dess and Beard (1984) even set the cut-off value for Cronbach's α to 0.6 for explorative studies.

Table 8. Cronbach's α (Cr. α), Composite Reliability (CR), Average Variance Extracted (AVE).

Cons.	Cr. α	CR	AVE
TAF	.908	.942	.845
COM	.970	.978	.918
EFF	1.000	1.000	1.000
INT	.952	.969	.912
BEN	.962	.972	.897
EXPL	1.000	1.000	1.000

The discriminant validity proves that items that measure different constructs are statistically different (Hair et al. 2013). To establish discriminant validity, we assessed the cross loadings as well as the square root of the AVE for each construct model (Fornell and Larcker 1981). As reported in Table 9 all constructs' square roots of the AVE are higher than their correlation to another construct. Due to the space restrictions we do not report the cross loadings, but we ensured that the loading of each item to its associated construct is greater than to other constructs. Thus, a satisfying convergent and discriminatory validity of the measurement model is given.

Table 9. Construct Correlations.

Cons.	TAF	COM	EFF	INT	BEN	EXPL
TAF	.919					
COM	.718	.958				
EFF	.582	.728	1.000			
INT	.567	.752	.657	.955		
BEN	.417	.413	.406	.359	.947	
EXPL	.624	.759	.644	.654	.470	1.000

We depict the results of the research model by running a bootstrapping with 5,000 re-samples (Davison and Hinkley 1997) in Figure 9. The model fit SRMR is .041, which refers to a good model fit since it is under the cut-off value of .08 (Hu and Bentler 1999). In H4a-e we postulated that advisor characteristics will positively affect the TAF. Only advisor's expertise affects the TAF (i.e., H4a is supported and H4b-e had to be rejected). Nevertheless, with this research model we can explain a high degree of variance in TAF as well as the advisor's expertise have a high effect size. The results of our research do not change by adding our control variables – age, gender, IT background, marker variable for common method bias.

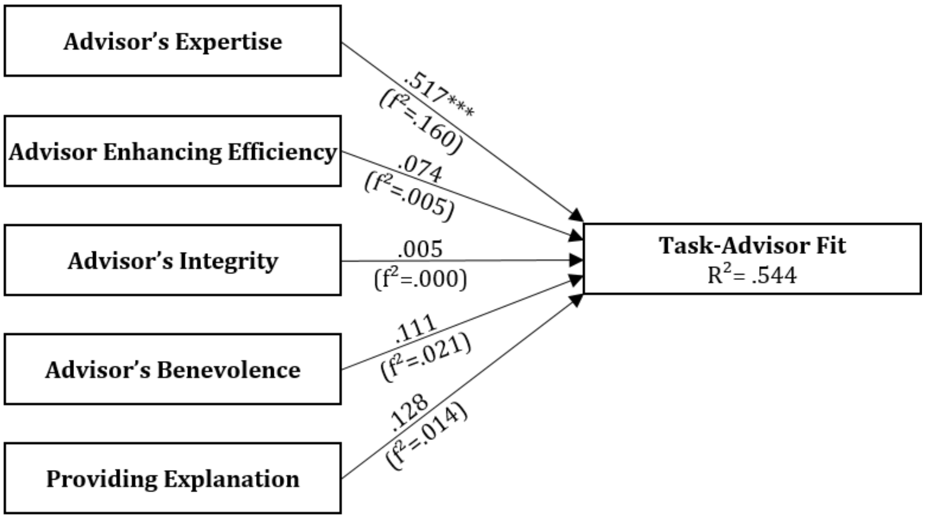

Figure 9. Result of Structural Model Testing.
*(*** p<0.001; ** p<0.01; * p<0.05).*

Discussion and contributions

The aim of our research was (1) to investigate whether there are differences in the utilization of advice from AI-based experts compared to human experts and (2) whether this is affected by the distance of decision-maker's initial estimation and advisor advice. Our research questions were addressed in an experimental study with 198 participants and in an online survey with 149 participants, thus contributing to the IS advice-giving and -taking literature.

Our main finding is that there are differences in the utilization of advice depending on whether it comes from an AI-based or human expert, which is also supported by the finding of Tauchert and Mesbah (2019) or by Logg et al. (2019). They also show this phenomenon but in other contexts like the financial ones.

The preference for AI-based experts in our experiment in comparison to human experts may be due to the participants' perception of a fit between advisor and task characteristics. It seems that primarily the competence of the expert plays a crucial role. The expert, who is generally assigned more competence for the task, appears to be preferred. As stated by Hoffmann and Krämer (2013), users prefer AI-based systems when a situation is task-oriented. Furthermore, the intention to use an AI-based system is greater when a user perceives a fit between technology and task characteristics (Chang 2010). By transferring this finding into the JAS literature, we show similar to Tauchert and Mesbah (2019) that the TAF reflects the advice utilization. Based on TAF we are able to evaluate if a preference between AI-based and human advisor exists for a specific task and which one is preferred.

For a better understanding of the nature of advice under which this phenomenon occurs, we have examined how the IS advice literature characterizes advice. An impactful characteristic is made by the distance between initial estimation and advice. Some researchers show that advices that are close to the initial estimation are more likely to be considered than far advices (Minson et al. 2011; Yaniv 2004a) However, there is also research with contrary findings, which conclude that more distant advice is given more weight (Schultze et al. 2015). Our study results support the second case, which states that the advice that is further away from the initial assessment of the participants is weighted higher than the closer advice. Looking at the group comparisons, however, it appears that AI-based advice was only preferred to human advice for the case of closer advice. According to Schultze et al. (2015) judges feel the need to adjust their estimation when receiving far advices based on the stimulus-response model. This need apparently leads to the fact that although AI-based experts are perceived more competent and are apparently preferred, in cases with far advice, these character differences between AI and humans are not strong enough to cause a difference in advice utilization. Summarizing, the answer to our research questions is that the advice of AI-based and human experts is used differently, but this effect is moderated by the distance of the advice.

Besides the theoretical contribution, our results have some practical implications. First of all, the results show that advices from AI-based experts are not necessarily discounted more than the advice from human experts. This allows professionals, depending on the task, to use AI-based advisors to automate processes and use the advantages of this technology. Secondly, the results show that providers of expert systems should use AI-based experts especially in situations where decision-makers themselves can estimate a situation well. This is derived from the insight that decision-makers are more likely to follow AI-based experts if their initial estimation is close to that of experts. Thirdly, a service provider can use the TAF to assess whether the implementation of an AI-based advisor is accepted. If the fit is not perceived as high as for human advisors the service provider is able to evaluate which characteristics influence this fit based on the task-advisor model and can influence and change the perception of these characteristics.

Limitation and future research

Certainly, there are also some limitations associated to our study. We compared the perception of AI-based and human advisors based on online experiments. That means participants have to imagine the situation of a real consultation. Certainly, the real interaction with a human or AI-based expert could lead to a different perception. Therefore, our findings should be validated in a more realistic laboratory experiment.

Another limitation is the simplification of the measurement model. In fact, the utilization of advice can be influenced by many different factors that can influence each other. In the following, some possible conditions and corresponding research questions for future research are presented.

Literature points out that previous experience and knowledge of users have influence on the intention to use a system (Rzepka and Berger 2018). Thus, expert systems should be preferred by users with little experience and knowledge for the given task (Will 1992). The resulting question would be whether decision-makers' previous experience and knowledge of the task have an impact on the utilization of AI-based and human expert advice.

Furthermore, we conducted a scenario-based experiment for one task only. Gino and Moore (2007). have proved that the degree of difficulty of tasks influences the extent to which the opinion of an expert is taken into account. The more difficult the task becomes, the more decision-makers take the opinion of an expert into account. An interesting aspect would therefore be to examine whether the level of difficulty of different tasks affects the preference of advisors.

5 Research Paper C: User-Group-Specific Acceptance and Utilization of AI-based Services

Title: "Hello, I'm Here to Help You" –
Medical Care Where It Is Needed Most: Seniors' Acceptance of Health Chatbots

Authors: Mesbah, Neda; Pumplun, Luisa

Published in: European Conference on Information Systems (ECIS), Online, 2020

Abstract

Demographic change is resulting in an increasing number of people getting older and needing adequate medical care. In order to solve this problem, easily accessible health chatbots could be established, which are capable of identifying diseases on the basis of symptoms. Such applications already exist today, but usage is primarily restricted to younger generations. Therefore, this study examines which factors affect the acceptance of health chatbots by seniors. With the help of 21 qualitative interviews within the respective target group and the extended Unified Theory of Acceptance and Use of Technology (UTAUT2) a comprehensive model for the adoption of health chatbots is developed. Additionally to being interviewed, the participants were able to test an exemplary health chatbot to better comprehend the technology. Thus, a practically oriented overview as well as 18 propositions were established, which could be used as a basis for further research regarding seniors' acceptance of health chatbots.

Keywords: Senior, Adoption, mHealth, Health Chatbot.

Introduction

The ongoing demographic change and the associated decline in comprehensive healthcare provision urgently require new developments in medical technology (Krys and Fuest 2017). A progress, that could help solve this social problem, can be found in the so-called mobile health (mHealth) area. M-health technologies "empower individuals to manage their personal health" (Kenny and Connolly 2017, p. 1129) and therefore make healthcare more efficient. Consequently, the market of mHealth is constantly growing. It is predicted to reach USD 151.57 billion by the year 2025 (Grand Review Research 2018). This is also reflected by the number of health apps available on iOS and Android. Over 100,000 apps with a medical focus can be downloaded to respective devices (Taylor 2015). A particularly promising progress in the mHealth area is the use of artificial intelligence (AI). Panetta (2018) state that AI-enhanced virtual care will help offer a more accessible, comfortable and cost-effective medical supply.

In this context, chatbots are a possible application form of AI. Chatbots are intelligent software programs that are able to communicate with users in text form by emulating natural language (Shawar et al. 2005). They are being used increasingly for health purposes, such as self-diagnosis. A particular chatbot is provided by Ada Health. Their app Ada works as a personal health guide and assesses symptoms by leading users through a directed dialogue while collecting relevant information. With the help of a medical knowledge base, patterns of symptoms are then examined for possible causes by applying intelligent algorithms (Ada Health 2019).

Especially elderly citizens can vastly benefit from such self-diagnosis tools based on AI. The term "elderly", as defined by the United Nations, includes all persons aged 60 years and older and is used accordingly in this study (United Nations 2017). This age group is particularly affected by chronic illnesses, such as osteoporosis, cardiovascular diseases or diabetes and therefore needs comprehensive medical care (Sanyal 2018). In order to facilitate the use of digital technologies by older people, much research is already being done concerning human-computer interaction (HCI) with the elderly as a target group. This target group has highly specific requirements and drivers regarding technology usage and adoption. Various factors, such as the perceived ability to live alone, play a major role concerning the adoption decision of the elderly (Vichitvanichphong et al. 2017). Unfortunately, according to Vichitvanichphong et al. (2017), older citizens are commonly resistant to changes.

In order to help overcome this obstacle and provide the elderly with sufficient medical care, the present study attempts to determine which specific factors influence the adoption of intelligent chatbots in healthcare (called health chatbots) by older people and integrate the findings into a clear overview. Our approach thus aims to identify:

Which factors influence the decision and ability of the elderly to adopt chatbots for self-diagnosis?

To answer this research question, an overview of related literature (i.e., chatbots in healthcare) is provided first. Afterwards, we present the theoretical background of our study, which we base on the extended Unified Theory of Acceptance and Use of Technology (UTAUT2). Afterwards, our study and the sample comprising 21 interviews with 23 interviewees is introduced. Our study is based on the application Ada, which can be used for self-diagnosis. Based on a first guided experience with Ada, participants are questioned and the empirical results are discussed and integrated into a model to reflect the specific needs of the elderly. The results of our paper provide a first overview of the age-specific factors (e.g., need for emotional support) that determine the adoption of chatbots in the healthcare sector. Thereby, our findings contribute to current research on the adoption of chatbots and show which factors must be taken into account, especially considering the continued progression of current demographic changes. Finally, we conclude our paper by pointing out possible limitations of the study and presenting directions for future research.

Related Research

Recent technological advances in the field of AI, machine learning and natural language processing (NLP) have renewed the interest of research and business in conversational interfaces (Seeger et al. 2018). Chatbots, as machine conversation systems that interact with human users through natural language, represent a specific application form of conversational interfaces (Shawar et al. 2005). Chatbots are text based systems that emulate human-to-human chat conversations and are therefore often perceived as anthropomorphic (Seeger et al. 2018). They can be differentiated in general-purpose or domain-specific bots according to their scope of application (Gnewuch et al. 2017). Possible specific purposes include the use in museums (e.g., Kopp et al. 2005), e-commerce (e.g., Qiu and Benbasat 2009), or in the healthcare sector (e.g., Minutolo et al. 2017).

Especially health chatbots (HC) have a long history, starting with Eliza, an early therapeutic chatbot based on a keyword search and stated rules (Weizenbaum 1966). In accordance with Wang and Siau (2018), we define HC as programs that are able to conduct intelligent conversations regarding health issues. Within the healthcare sector, chatbots can help solve a variety of problems. For example, they may be used to recommend preventive health measures (Amato et al. 2017). However, they can also assist in the follow-up care of patients by reminding users to take their medication, tracking users' health trend or finding the nearest pharmacy (Wang and Siau 2018). In this paper, we focus on HC which are able to identify possible diseases based on entered symptoms (Minutolo et al. 2017). In a society facing demographic change and the obligation to deal with high life-expectancies as well as urbanization, medical supply is becoming increasingly costly (Calero Valdez and Ziefle 2019). Medical processes must therefore increase in efficiency. One solution could be to examine patients by intelligent machines in advance and decide which diseases require medical treatment (Hoermann et al. 2017). This allows clinicians to spend more time on real emergencies (Bibault et al. 2019). Therefore, the usage of self-diagnosis HC could "likely become the first point of contact for primary care" (Wang and Siau 2018, p. 1). Real products like Ada, Babylon Health or Your.Md prove a first feasibility of such systems (Ada Health 2019; Babylon 2019; Your.MD 2019). Nevertheless, various factors could affect the extensive adoption of HC. First, the functionality of intelligent chatbots relies heavily on provided data. Therefore, the quality as well as privacy of this data is crucial (Wang and Siau 2018). Another aspect could be the lack of trust in the recommendations of HC that could arise from communicating with a dehumanized entity (Amato et al. 2017). Additionally, Fadhil and Schiavo (2019) state that HC should take into account users' demographics. Although these factors clearly influence users' decision of adoption, little is known about the respective needs of different user groups interacting with HC. To the best of our knowledge, there is only one contribution that considers age specific adoption factors of HC from a more theoretical perspective and integrates them into a comprehensive model. This study concentrates on students with a mean age of 24.8 years. This age group may more likely belong to the group of early adopters, but has a much lower need for and easier access to medical care than the elderly. Therefore, the authors call for further studies, in which senior citizens are considered (Laumer et al. 2019).

In order to meet this demand, we are building our study regarding the technology adoption of elderly people on the extended Unified Theory of Acceptance and Use of Technology for the consumer context (UTAUT2). UTAUT2 is the most recently developed and discussed model to study technology acceptance from a consumer's point of view. As we investigate seniors' acceptance regarding HC, we consider UTAUT2 as particularly helpful in order to examine the individual adoption of a specific consumer group, i.e., the elderly (Venkatesh et al. 2012). Moreover, we chose to use UTAUT2 as the underlying model for our research as it was used in pertinent contexts, namely healthcare and seniors' technology acceptance (Chen et al. 2014; Laumer et al. 2019). In the following, we thus explain the UTAUT2 model in more detail (see Figure 10).

Performance expectancy is "the degree to which using a technology will provide benefits to consumers in performing certain activities" (Venkatesh et al. 2012, p. 159), whereas *effort expectancy* is defined as "the degree of ease associated with consumers' use of technology" (Venkatesh et al. 2012, p. 159) – or in other words: how easy is it for someone to use the technology studied? *Social influence* describes the impact of people important to the deciding person (e.g., family, friends) on their technology usage. *Facilitating conditions* "refer to consumers' perceptions of the resources and support available to perform a behavior" (Venkatesh et al. 2012, p. 159). *Hedonic motivation* refers to the enjoyment or pleasure a person receives from using a technology, whereas the *price value* represents the result of the cognitive trade-off between the perceived benefits and monetary costs of a technology. *Habit* is a construct that indicates the extent to which the beliefs and behavior of a consumer is described as automatic, since they have become a routine over time. The factors described influence the *behavioral intention* to use a technology, which serves in turn as a key predictor of the actual *use behavior*. The dependent variable *use behavior* is furthermore directly affected by the facilitating conditions and habit (Venkatesh et al. 2012).

The basic UTAUT2 model describes different factors that determine the acceptance and use of technologies by consumers in general. Since HC are applied in a highly sensitive context (i.e., healthcare) and require a very specific and new use behaviour compared to other technologies, it is crucial for researchers to reflect on these particular artefacts (Davison and Martinsons 2015). Moreover, various studies have shown that seniors pose their very own needs in terms of complex technologies (Vichitvanichphong et al. 2017). Both, the novelty of technology and specific requirements of seniors make it necessary to review, adapt, and extend the basic model (i.e., UTAUT2) according to the specific research context.

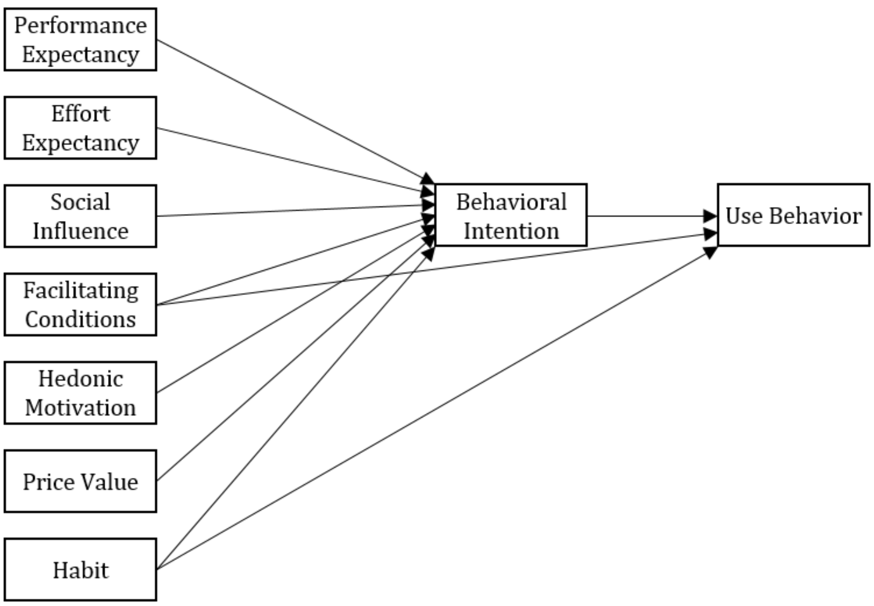

Figure 10. UTAUT2 Model.
(Based on Venkatesh et al. 2012).

Qualitative Research Methodology

The aim of the study is to close the current research gap regarding the acceptance of HC by older people. Since the use of HC is a new, little researched topic, an explorative approach was chosen in which the target group of seniors is interviewed qualitatively (Flick 2004). To do so, a three-stage qualitative approach was used (Elliott and Timulak 2015). At the beginning, the current state of literature was searched with regard to technology acceptance of older people as well as adoption in a HC context (see Section 2). Based on the achieved conceptual foundation and the principles of Sarker et al. (2013), a semi-structured interview guideline was established, which was used to conduct in-depth qualitative interviews with participants. The research process was concluded by transcribing, coding and analyzing the interviews with the aim of showing senior-specific adoption factors for HC.

Research Design, Sample, and Data Collection

The semi-structured interview process comprised three different phases which were accompanied by the interview guide. The first phase included general questions on demographic factors, personal technology affinity, and health conditions. During the second phase, participants were encouraged to use the Ada app, which served as a representative example to test the handling of HC. While using the application, participants were provided help if necessary and were animated to share their experiences for the recording. By actually using the app, it was ensured that participants could better assess the advantages and disadvantages of HC. In the last and most comprehensive part of the interview process, questions about the acceptance of such chatbots were asked. Only open questions were demanded (e.g., "Would you use Ada and why?") in order to give the participants sufficient freedom to describe specific acceptance factors not yet included in UTAUT2. Due to the nature of the semi-structured approach, interviewers were able to make further inquiries and take up new aspects mentioned by the participants (Bryman 2016; Myers 2013). In this context, we used the laddering technique to ask successive "why" questions (Reynolds and Olson 2001).

The qualitative data was collected over a two-week period and took place in May 2019. This timeframe was chosen to ensure that all participants used a consistent version of the app. In total 21 interviews with 23 participants from our target group "senior citizens" (60+ years) were conducted within Germany. Among the seniors were 12 female and 11 male participants, whose ages were between 60 and 96 years (mean = 71 years). 20 respondents were already retired, while 3 participants were still working. Half of the participants would describe themselves as technology affine, although none of them had used a chatbot before. Nevertheless, the relevance of HC was high for the selected sample, as many of the participants already had health restrictions and were therefore forced to see the physician frequently. After the 21st interview, data collection was discontinued as only redundant aspects arose in the interviews (i.e., theoretical saturation was assumed) (Flick 2004). The interviews lasted an average of 38 minutes and were exclusively conducted face-to-face to enable the active use of the app and to take into account the complexity and sensitivity of the topic. The interviews took place in closed rooms, guaranteeing freedom from disturbance and external influences. All participants took part in the interviews on a voluntary, non-paid basis and have been assured of their anonymity.

Content Analysis

As recommended by Weber (1990), the assessment of interviews was based on content analysis, which is particularly applicable to the investigation of open-ended questions. Following the steps of content analysis, UTAUT2 was used as a foundation to conduct the study. To include both the known factors as well as new aspects mentioned by seniors, a combination of directed and conventional analysis was used. While the directed approach accounts for constructs known from relevant literature (i.e., UTAUT2), conventional analysis enables to critically evaluate, extend or subdivide initial factors based on new findings from the qualitative data (Hsieh and Shannon 2005). To make this information available, all interviews were recorded in consultation with participants and transcribed immediately after conducting the interviews in order to ensure that no relevant content was lost. Afterwards, transcripts were coded using the NVivo 12 software. As recommended by Saldaña (2009), coding was conducted via two constitutive coding cycles. The first cycle included a combination of attribute, hypothesis and descriptive coding. Attribute coding was conducted to gain insights about demographic data (e.g., participant's age, gender). The hypothesis coding took into account the initial factors of UTAUT2. Afterwards, descriptive coding was used to summarize relevant passages of the derived qualitative data, which were specifically reflecting the needs of seniors in context of HC. During the second coding cycle, pattern coding was used to evaluate the previously generated codes critically and combine them into a smaller number of categories. In the sense of an investigator triangulation, achieved factors were discussed in a group of four Information Systems (IS) researchers and students.

Results and Discussion

Within our analysis, we were able to confirm the applicability of the UTAUT2. Nevertheless, the factors as defined in the seminal work by Venkatesh et al. (2012) are not sufficient to explain the acceptance and use of HC by seniors. Therefore, we have extended the model by redefining existing factors, adding new ones, and excluding factors which could not be identified as relevant within the interviewing process (i.e., hedonic motivation). Among the most mentioned new factors were the need for emotional support (14 of 21 interviews), technology self-efficacy (13 of 21 interviews) and medical history (11 of 21 interviews). Furthermore, the relevance of the factor price value could not be examined based on the Ada app, as it is made available to users free of charge. The result of our study is shown in Figure 11. In the following we will explain and discuss our findings in more detail.

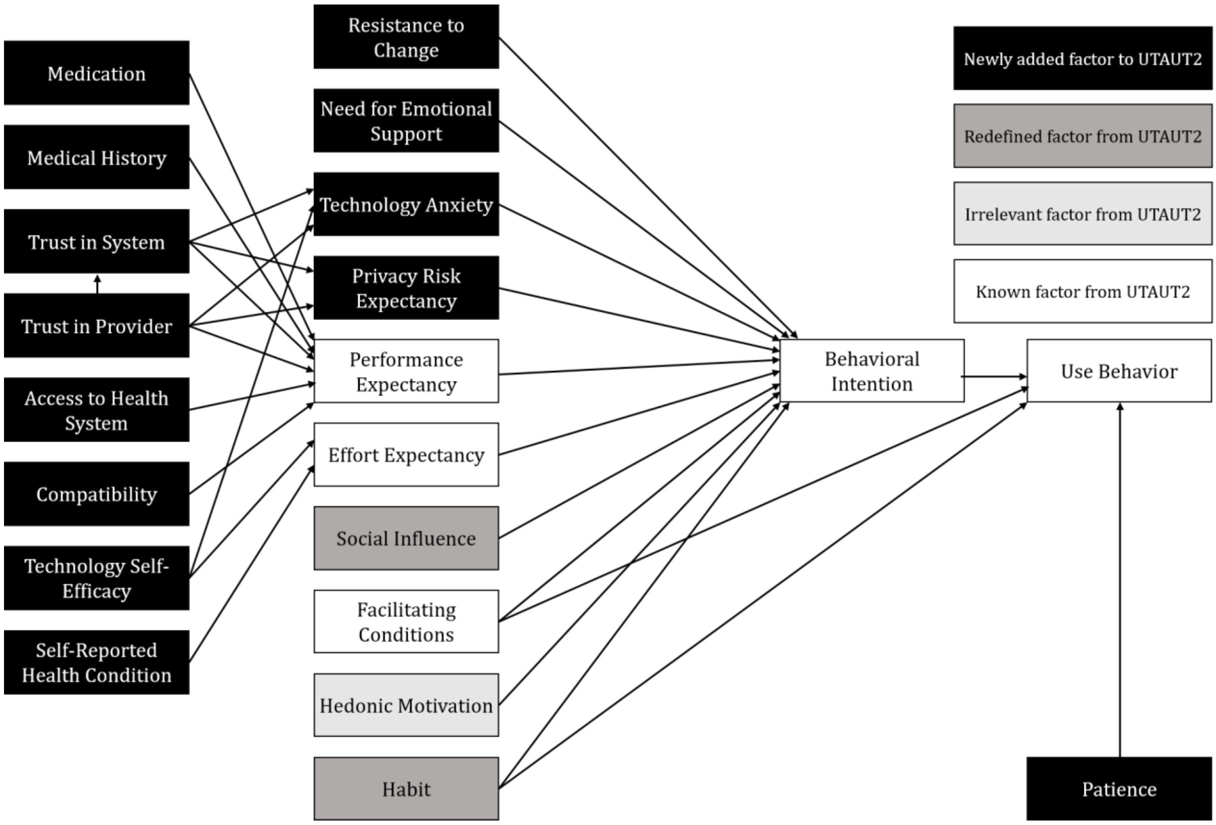

Figure 11. Seniors' Acceptance of HC – an Extended UTAUT2 Model.

Known UTAUT2 Factors

In the following, the factors adopted from the UTAUT2 model are discussed in more detail.

Performance Expectancy

In line with existing literature on technology acceptance, the interviews have shown that the usage intention of seniors is strongly dependent on the *performance expectancy* regarding the HC. Seniors said that the usage of HC is time-saving compared to visiting the physician and allows them to quickly obtain information to assess how urgent a visit to the physician is or to alleviate worries of a serious condition. However, it was further stated that from their point of view, this cannot replace the first visit to the physician. In the eyes of the elderly, HC are a kind of supporting tool in case a physician is not immediately available or to follow up their past visit to the doctor. In this context, seniors would rather use the tool to gain more knowledge about the diagnosed disease and to use this information as a basis for further conversations with the physician. This is illustrated by the following statement:

"Also to be able to understand the statement of the physician is quite nice [...], because the visit at the physician happens quickly in a relatively short time. It is not necessarily always comprehensible […]. For this it [i.e., Ada] can help very well." #5

Another aspect that is part of the performance expectancy of seniors towards HC is the ability of the chatbot to respond to their individual needs, similar to what a physician would have done:

"No, I don't want to be informed about a machine first. I want to see a physician. [...] Because it [i.e., Ada] should be more individualized, but I don't know how that would work. As I said, because there are so many different people out there." #11

There is a wide variety of research in the field of HCI that shows that technology should fit to the individual needs of seniors in order to encourage a usage intention (e.g., Kim et al. 2018): For example, their medical demands change over time and technology should be flexible enough to adhere to these new circumstances (Gao and Koronios 2010). Hence, we have formulated the following proposition (**P1**):

The higher seniors' performance expectancy, the higher is seniors' intention to use HC.

Effort Expectancy

Another crucial factor confirmed by the interviews is *effort expectancy*. It is crucial for seniors to be able to use the HC with little effort. For example, the handling of the HC should be preferably simple and the answer design within the dialogue should be sufficient, precise, and easy to understand. In this regard, some of the possible answers and explanations of the respective HC were not comprehensive enough for the participants to be able to understand their intention and to answer appropriately, as can be seen from the following comment:

"The explanation was very poor. For example, 'at the hollow of the knee'. I wanted to know if I understood that correctly. The hollow of the knee is what is directly below the knee, isn't it? [...] And then he just told me again, 'yes, pain in the hollow of the knee'. But I would have liked [...] to receive more information." #12

All in all, the use should be simple and answers should be easily understandable for seniors in order to increase their willingness to use the HC. Therefore, our second proposition is (**P2**):

The higher seniors' effort expectancy, the higher is seniors' intention to use HC.

Facilitating Conditions

Another challenge to accept and use HC, which is critical for seniors, is to own a smartphone and have access to the internet. These requirements were mentioned by several interviewees, showing the relevance of this topic for seniors. Therefore, *facilitating conditions*, as already shown in the literature (Venkatesh et al. 2012), have a particularly high relevance for the usage intention and behavior of seniors regarding HC. However, the technological requirements are not the only challenge that seniors face when intending or actually using HC. This segment of the population often needs appropriate and long-term assistance in order to use digital systems and reap their full benefits (Chen et al. 2014). This is particularly the case when it comes to novel and complex technologies, such as HC. With the assistance of other people, seniors feel more confident to accept and finally use an HC:

“So now, as you are here and explain it to me, I think it's easy. But I don't [...] know what it would be like, if I was left on my own.” #4

Based on the pertinent literature and our interviews we thus state (**P3**):

a: The better seniors’ facilitating conditions, the higher is seniors’ intention to use HC.

b: The better seniors’ facilitating conditions, the higher is seniors’ actual use of HC.

Redefined UTAUT2 Factors

The factors social influence and habit known from the original UTAUT2 had to be slightly redefined.

Social Influence

In agreement with the research on technology acceptance, our interviews show that the social environment has a large influence on seniors’ acceptance of HC. So far, *social influence* has been interpreted as consumer’s perception of how much the social environment believes a technology should be used by the consumer and demands her or him to do so (Venkatesh et al. 2012). However, our interviews show that besides the active call of others to use a technology, the personal assessment and recommendation of the social environment play a crucial role for seniors’ intention to use an HC as can be seen from this statement:

“It's a basis of trust when you say it's great. […] And then it [i.e., Ada] will be accepted.” #13

We thus redefine the factor in accordance with Laumer et al. (2019) “as the extent to which consumers perceive that important others believe they should use a particular technology, that important others say they should use a particular technology, that important others recommend to use a technology or that individuals observes that others are using a specific technology” (Laumer et al. 2019, p. 7).

There are several other studies which show that social influence plays a key role for seniors’ technology acceptance (e.g., Guan et al. 2017). Based on the new definition of the construct, we conclude our third proposition (**P4**):

The higher seniors’ social influence, the higher is seniors’ intention to use HC.

Habit

In the interviews, it turned out that none of the seniors has knowingly used a chatbot so far. Therefore, in contrast to Laumer et al. (2019), we found no support for the proposition that the *habit* to use chatbots in general, has an influence on the acceptance and use of HC in particular. Instead, it has become apparent that many seniors are used to visit the physician in person (e.g., because of chronic illnesses) and unlike the younger generation do not have the habit to search for their symptoms online in order to diagnose themselves. Although the usage of HC is quite new for younger persons, too, they have grown up using current technologies such as search engines. Therefore, the first step towards using a machine to assess illnesses is less unfamiliar for younger people than it is for older ones:

"The challenge is even if you're technology affine or have a good knowledge of digital media, but you've been visiting a physician for 60, 70 years [...] it's a challenge to not deal with a human." #12

Our results confirm the findings of researchers that seniors are struggling to adapt to the rapid changes entailed by current technologies and to alter their habits accordingly (Holgersson and Söderström 2019). We thus redefined habit as the habit of seniors to visit physicians instead of searching online or consulting HC, resulting in the following proposition (**P5**):

a: The higher seniors' habit of visiting physicians, the lower seniors' intention to use HC.

b: The higher seniors' habit of visiting physicians, the lower seniors' actual use of HC.

Newly Added Factors

In addition to the original factors, the interviewees also addressed factors that are specific to the seniors' acceptance and use of HC. These factors will be explained below.

Patience

The interviews revealed *patience* of the seniors as a factor that influences seniors' usage behavior directly. As already described, the handling of HC is new for seniors and demands that they have patience while learning to use the technology. Accordingly, dealing with HC requires more patience from seniors than from younger generations, who are more used to deal with related technologies such as chatbots in general. However, seniors often do not have the patience to learn how to use a new app (Carlsson and Walden 2015). The inherent characteristic of HC to guide their users through a dialogue with several questions and query loops about symptoms and health conditions amplifies this problem further:

"When you forgot to enter an answer during the usage of the application, maybe you can't enter it anymore, then you get confused and impatient." #11

In other words, the patience of seniors with HC plays a crucial role in the final usage or rejection of HC. Accordingly, we propose (**P6**):

The higher seniors' patience, the higher seniors' actual use of HC.

Resistance to Change

Since senior citizens mostly have limited experience with technologies, even small adjustments of known app design and user interfaces can cause major defensive reactions. In this context, the inherent characteristic of HC to guide the user through a longer dialogue has led to difficulties for the seniors. For example, the navigation of the HC Ada was different from other known apps, which makes handling less intuitive for them. In this regard, in order to return to the previous page, participants have to click on the greyed answer to the earlier question instead of using a familiar back button. This *resistance to change* can be explained by resorting to a theory known from psychology. According to the theory of fluid and crystallized intelligence, humans own two kinds of intelligence: i.e., fluid and crystallized intelligence. Crystallized intelligence refers to abilities gained from experience, whereas fluid intelligence describes whether people are able to adapt and solve new situations as they occur. With age, crystallized intelligence increases while fluid intelligence decreases (Cattel 1963). However, this makes it more difficult for seniors to adapt to unknown procedures. Consequently, seniors are more resistant to changes (Hoque and Sorwar 2017). As a result, we formulated the following proposition (**P7**):

The higher seniors' resistance to change, the lower seniors' intention to use HC.

Need for Emotional Support

During the interviews it was most frequently mentioned that the diagnosis by a machine as provided by the HC is not sufficient for seniors. The participants emphasized the importance of a human who knows you, has cared for you for years, who you are familiar with, who you trust, with whom you can have a dialogue and who you can ask questions if necessary. This is pointed out in the following statements:

"But at the physician's, I could also ask more in-between questions. And the whole process [in Ada] is planned, or rather pre-set, and the app doesn't allow interposed questions." #23

"But I think it's the personal relationship. This feeling, you are taken seriously, that you are valued, that people listen to you. The other person is trying, to get to the root of your pain or complaints and find out more. […] And this patient-physician relationship, that is also unique, I think. So in my case […] my physician, who made the first diagnosis breast cancer, he conveyed that we will get it all done [...] And that's what I mean when I am speaking about personal relationship. You don't have to die of cancer anymore. We can manage that. And because of this charisma of the physician, I'm sure that was a giant healing factor for me.” #8

The *need for emotional support* can be of great relevance for seniors in particular, since they often have less social contact and more severe illnesses than the younger generation (Wilson 2018). In particular, the dialogue with the HC cannot replace the need for human empathy. Therefore, we conclude the following proposition (**P8**):

The higher seniors’ need for emotional support, the lower is seniors’ intention to use HC.

Technology Anxiety

Few interviews have shown that working with HC is a big and daunting challenge for seniors due to their lack of digital literacy. Seniors did not grow up with technology in their everyday lives. Rather, they were confronted with it at an advanced age. As a result, they have a lower level of technological experience compared to the younger generation. This leads to greater skepticism and fear regarding technology in general, as can be seen in the following:

“But I just think [...] that many people are being cheated and fooled through the internet and modern forms of communication. So we have to be extremely careful with everything we do, right?” #15

Our interviews have shown that among seniors not only the fear of technology itself, but also the fear of making mistakes while using it is severe. Furthermore, seniors state the anxiety of being cheated through a technological application. Hence, we call *technology anxiety* not only the fear of using technologies, but also the fear of being deceived while using it. HC, which are based on complex AI and NLP algorithms, can be judged even less by senior citizens than conventional technologies. Therefore, technology anxiety significantly affects the intention of seniors to use HC. In this context our ninth proposition is defined as follows (**P9**):

The higher seniors’ technology anxiety, the lower is seniors’ intention to use HC.

Privacy Risk Expectancy

The interviews have shown that *privacy risks* and data protection regarding the entered symptoms are a factor which seniors take into account. For example, the following consideration is expressed during the interviews:

“Who can guarantee me that I can work with this system safely? On the one hand, I know it cannot work, if I do not give qualified information. On the other hand, I would be worried depending on the type of the disease. I do not necessarily want everyone to know it.” #19

However, opinions on privacy were controversial among the participants. In this regard, a lot of participants did not consider the protection of data as necessary and preferred a well-founded diagnosis:

“My protected data will no longer be useful to me if I am no longer alive.” #13

“Anybody can hack my data if they want, I don't care. For me, a well-founded diagnosis is more important. That's what I would use this app for, because I need information, a well-founded diagnosis.” #3

Compared to the younger generation, privacy risks could not be identified as a decisive factor (Laumer et al. 2019). We thus conclude that the importance of a well-founded diagnosis becomes higher with age and therefore exceeds the fear of privacy risks. Nevertheless, seniors tend to consider privacy issues as relevant. Based on these insights, we have concluded the following propositions (**P10**):

a: The higher seniors’ privacy risks expectancy, the lower is seniors’ intention to use HC.

b: The higher the need for an informed diagnosis, the lower is the negative impact of privacy risks expectancy on seniors’ intention to use HC.

Medication

Many seniors need medication to treat chronic diseases (Charlesworth et al. 2015). Thus, participants expressed the desire that Ada should consider prior *medication* and potential side effects to achieve a well-founded diagnosis. However, the current version of Ada is not able to consider pre-medication and their potential effects on health for diagnosis. Furthermore, a few seniors mentioned that they need to see a physician to get new prescriptions for medication. Since these seniors, who need medical treatment, visit the physician anyway, the perceived performance expectancy of HC decreases:

“I wouldn't get any medication through this app [Ada] [...].” #3

Due to the current technological limitations, we have formulated the following proposition (**P11**):

The higher the amount of seniors’ requirement for medication, the lower is seniors’ performance expectancy of HC.

Medical History

The older people get, the longer their *medical history* is and the more likely they are to already suffer from a disease (Charlesworth et al. 2015). This history of chronic illnesses can have a crucial influence on current symptoms entered to the HC. Therefore, the use of HC only makes sense if they take into account the individual anamnesis of seniors, as the following statement shows:

“But for my case, it doesn't fit. [...] I had a stroke. Naturally I have strange sensations and things that feel like pain or cold or something and it [the answer of the HC] doesn't fit right.” #7

Due to their longer lifespan, seniors are more likely to have had contact with the topic of diseases and medical treatment (Charlesworth et al. 2015; World Health Organization 2019). Therefore, more than half of the interviews have shown, that seniors' performance expectancy is strongly dependent on the complexity of their diseases:

“She [Ada] can identify simple symptoms and diagnose them correctly, but as it becomes complex, I would no longer trust Ada, because there can be many different causes for one symptom.” #8

The interviews showed that the medical issues of seniors cannot be neglected and that they are decisive for the performance expectancy of HC. As a result, we conclude the following proposition (**P12**):

The larger seniors’ medical history, the lower is seniors’ performance expectancy of HC.

Trust in System

As in the previous study on HC (Laumer et al. 2019) the interviews showed that *trust in system* has an influence on the extent to which the HC is perceived as privacy friendly, how skeptical the seniors are towards the technology, and how the performance of the HC is expected to be. In order to increase the trust of seniors in HC, it is necessary to provide them the possibility to test the system. Interviewed seniors have not yet gained any experience with HC. Therefore, in many interviews participants stated that they could only build trust by using the system and checking whether it provides correct diagnoses. This is stated in the following example:

Well, I'd always have it cleared up, I guess. If she [Ada] now tells me, for example, that you have osteoarthritis, [...] then I would go to the physician first and have it checked. And if he also says that I have osteoarthritis, then I would trust the system more. Next time, if I [...] have diarrhea or something and she gives me a diagnosis, then maybe I would say: Last time, she was right. So maybe it's okay this time, so it's right.” #8

Accordingly, we propose (**P13**):

a: The higher seniors’ trust in system, the higher is seniors’ performance expectancy of HC.

b: The higher seniors’ trust in system, the lower is seniors’ technology anxiety of HC.

c: The higher seniors’ trust in system, the lower is seniors’ privacy risk expectancy of HC.

Trust in Provider

As shown for younger age groups, *trust in provider,* similar to trust in system, is a key factor in the adoption decision of the elderly (Laumer et al. 2019). In the case of seniors' usage, however, it was mainly important to the participants that the HC is supported and co-developed by physicians and that the data is generated and approved by physicians. This can be seen in the following statement:

"Yeah, well, is it written by physicians? Do physicians stand behind it?" #11

The high relevance of this factor can be explained by the fact that this age group has a very high respect for the profession and the knowledge of doctors (Marcinowicz et al. 2014). Therefore, seniors call for transparent explanations of provider's product development process of the HC and of the creation of the utilized database. The interviews have also shown that trust in the provider increases seniors' performance expectancy towards the HC, lowers the perception of privacy risks and technology anxiety, and builds trust in the system, and. Therefore, we propose (**P14**):

a: The higher seniors' trust in provider, the higher is seniors' performance expectancy of HC.

b: The higher seniors' trust in provider, the lower is seniors' privacy risk expectancy of HC.

c: The higher seniors' trust in provider, the lower is seniors' technology anxiety regarding HC.

d: The higher seniors' trust in provider, the higher is seniors' trust in the system.

Access to Health System

Access to public healthcare services is limited for many people, especially the elderly (e.g., Steele et al. 2009). This is the case as seniors are often not mobile enough to autonomously access public healthcare services. Furthermore, the physicians are often very busy and thus hard to contact. Technologies like HC can partly counteract this problem. In nearly half of our interviews various scenarios are described, where access to a medical professional was difficult:

"The physicians are overloaded. You can't reach the physicians, sometimes for days." #11

HC could make it easier and faster for seniors to get access to medical advice, increasing their performance expectancy regarding this technology. Thus, we conclude the following proposition (**P15**):

The lower seniors' perceived access to health systems, the higher is seniors' performance expectancy of HC.

Compatibility

Our interviews have shown that HC hold the potential to serve as a central repository for medical records (as symptoms are entered regularly) and as a shortcut to a wide variety of medical knowledge (as it is based on a growing database). If HC would be integrated in the healthcare system, healthcare processes thus could be optimized, for example by avoiding unnecessary repetitive examinations, and seniors' performance expectancy would increase. This is illustrated by the following statement:

"When you can simply say: everything [diagnoses and test results] is available and the physician […] can access all necessary information just by a click, I would always use it immediately." #21

Accordingly, we have concluded the following propositions (**P16**):

The higher the compatibility of HC with seniors' healthcare system, the higher is seniors' performance expectancy of HC.

Technology Self-Efficacy

During the interviews the seniors very often felt that they do not have the competences to make use of the HC properly on their own. This had an influence on seniors' perception of how much effort the technology usage requires. This aspect is illustrated in the following statement:

"I somehow have no understanding for the technology, I have become more cumbersome" #14

In this context, Chen et al. (2014) also have shown that an increased technology self-efficacy has a positive impact on seniors' perceived ease of use regarding gerontechnology. As a result, *technology self-efficacy* has to be increased in order to increase the seniors' effort expectancy. In this regard, studies have shown that trainings are helpful to introduce seniors to information and communication technologies (Nguyen et al. 2014). Furthermore, the interviews have shown that more technologically competent seniors feel less anxious in dealing with technology. As a result, we formulate the following propositions (**P17**):

a: The higher seniors' technology self-efficacy, the higher is seniors' effort expectancy of HC.

b: The higher seniors' technology self-efficacy, the lower is seniors' technology anxiety regarding HC.

Self-Rreported Health Condition

The interviews showed that the more seniors were restricted in their health, for example in terms of visual or auditory ability, the harder it was for them to use the HC. The restrictive effect of a poor *self-reported health condition* on effort expectancy can also be confirmed by the findings from other senior technology acceptance research (Chen et al. 2014). As a consequence, it is especially important to seniors to have an age-appropriate visualization of the HC. In this regard, it was mentioned several times that the font size as well as the distance between the buttons of Ada were too small and that the visual contrast of the app should be higher. This can easily be explained by the fact that eyesight decreases with age (Glasser and Campbell 1998). Therefore, we proposed (**P18**):

The better seniors' self-reported health condition, the higher is seniors' effort expectancy of HC.

Conclusion, Limitations, and Future Research

In summary, we conducted an exploratory qualitative study by interviewing 23 participants in 21 interviews with a mean age of 71 years. This procedure allowed us to develop senior-specific factors regarding the acceptance and use of HC. As a result, we were able to confirm the applicability of the UTAUT2 model as a theoretical foundation. However, besides the already known factors of UTAUT2, the interviews showed that two factors had to be redefined and thirteen new factors had to be added. Furthermore, it turned out that the factor hedonic motivation did not seem to be relevant in the context of seniors' acceptance and use of HC. Based on our extended model it is now possible to evaluate the factors affecting the usage intention and subsequent behavior of seniors regarding HC in order to be able to fully exploit the advantages of HC for elderly people.

Although a study of Laumer et al. (2019) has identified factors specific to the acceptance and use of HC by the younger generation before, it cannot be assumed that the results are transferable to the context of seniors as this target group has highly specific requirements for the use of technologies. Furthermore, existing research on technology acceptance of seniors is also not sufficient to explain the intention and actual use of HC by older people, as HC is a novel technology in a highly sensitive context (i.e., healthcare) that was not investigated in detail before. Therefore, we contribute to the IS research of seniors' technology acceptance theoretically by identifying senior-specific factors that influence the intention and usage behavior of this specific mHealth service. In general, our results show that the participants not only showed a high need for technological assistance in order to use HC, but also expressed a greater demand for emotional support than younger people. Furthermore, seniors might have problems using currently available HC solutions with regard to their medication or medical history, which are not yet considered adequately by HC. In addition, it cannot be assumed that seniors are used to handle novel and complex chatbot technologies and therefore have issues trusting them.

Beside these general findings, the relevance of some factors is particularly noteworthy. First of all, the interviews illustrated the relevance of context-specific factors (e.g., medication, medical history) that should be considered for HC solutions used by elderly people. This finding could also be transferred to technologies other than HC (e.g., intelligent pillbox, wearable medical devices) in order to increase seniors' performance expectancy of the respective technology (Solís et al. 2017; Zhang and Li 2017). Among all identified context-specific factors, need for emotional support should be particularly emphasized. When it comes to seniors' health issues, emotions play a major role. In this context, studies have shown that a positive attitude can promote recovery (Glass and Maddox 1992; Shyu et al. 2006; Tsouna-Hadjis et al. 2000). Especially for seniors, who have a higher risk for serious illnesses and a higher probability of being less socially involved than the younger generation, emotional support is crucially important to deal with serious diagnoses. In particular, the support of a physician, they have known and trusted for many years, can be required in the event of a serious illness. As a consequence, the fact that machines do not offer emotional support, but rather perform very rational, short and concise diagnoses, can seriously hinder the acceptance and use of HC depending on the health condition of the senior citizen. Therefore, the questions arising are, for which types of diseases, according symptoms, and sub-groups of seniors the use of HC makes sense and how should the diagnoses of HC be designed in order to give users a feeling of emotional support (e.g., appropriate presenting of diagnosis, digital follow-up discussion with physician, references to help-offerings)? Given the high relevance of the factor need for emotional support, future research could furthermore investigate whether all identified factors influencing the acceptance and usage of HC should be equally prioritized and how context and specific environmental situations could affect this weighting.

Our study also offers some practical implications. In order to leverage the high potential of HC for the elderly, HC providers should design their offerings according to the specific needs of the elderly. For example, in order to ensure seniors' acceptance and continuous use of HC, it is essential that their medication and medical history is taken into account when they are guided through the dialogue with the machine. Since older people in particular are more likely to have pre-existing conditions, consideration of medication and medical history is particularly important to ensure a higher performance expectancy exhibited by seniors. In addition, the visual design of HC should be kept as simple as possible, based on known procedures, and appropriate for persons with weaker eyesight. Furthermore, the answer design should be easy to understand, but comprehensive enough in case of inquiries. Finally, as seniors need appropriate facilitating conditions and an enhanced technology self-efficacy to accept and use HC, procedures should be created to explain the advantages and usage of HC step by step. This could be done by introductory descriptions within the system itself or through trainings in cooperation with insurances and physicians.

Due to the explorative nature of this study, only 21 interviews are considered, which could result in a potential generalizability problem. However, we applied various criteria to ensure rigor and trustworthiness to encounter this problem (i.e., theoretical saturation, investigator triangulation).

In addition, the study was only conducted within one European country (i.e., Germany). In order to examine possible cultural influences, further studies should be executed in other countries and continents. In addition, the context-specific factors identified within this study could be investigated in more detail. In this regard, researchers could examine in which situations and under what conditions HC are used by seniors (e.g., for which symptoms). The interviews showed that there is a general willingness to use HC among seniors, but it does not replace a visit to a physician for them. Therefore, future studies should consider that HC will be used in parallel with a visit to the doctor and represent an additional offer for senior citizens. Overall, we have established a model that identifies the factors that contribute to senior citizens' acceptance and use of HC. This model could serve as a basis for further research, e.g., concerning seniors' acceptance of other related AI technologies (e.g., humanoid robots).

6 Research Paper D: Promotion of AI-based Services

Title: Promoting Trust in AI-based Expert Systems

Authors: Mesbah, Neda; Tauchert, Christoph; Olt, Christian M.; Buxmann, Peter

Published in: Americas Conference on Information Systems (AMCIS), Cancun, Mexico, 2019

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.

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 2016b). 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.

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, p. 2; McCarthy 2007). 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 (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.

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 2016b; 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 Appendix 2. 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.

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 10). 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, $\chi2(5) = 83.944$, $p = .000$. 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 $\eta2 = .064$.

Table 10. Mean and Standard Derivation of Relative Advantage Constructs.

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

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 11. The mechanisms can be sorted as follows: *Testing, Data Transparency, Reasoning, History, Training Frequency, Confidence, Usage Time, Social Environment, Technical Functionality, Dialog, Visual Appearance*.

Table 11. Mean, Standard Derivation and Rank of Trust-Increasing Mechanisms.

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

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.

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). 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.

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 and Weinhardt 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.

Acknowledgement

This project was funded by “Hessisches Ministerium des Innern und für Sport”.

7 Research Paper E: AI-based Services in Organizations

Title: Coordinating Human and Machine Learning for Effective Organizational Learning

Authors: Sturm, Timo[a]; Gerlach, Jin P.[a]; Pumplun, Luisa[b]; Mesbah, Neda[b]; Peters, Felix[b]; Tauchert, Christoph[b]; Nan, Ning[c]; Buxmann, Peter[c]

Authors with the same superscript contributed equally to this paper

Published in: MIS Quarterly 2021, 45 (3), S. 1581-1602

Abstract

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

Keywords: Artificial Intelligence, Machine Learning, Human-Machine Coordination, Organizational Learning, Simulation, Agent-Based Modeling.

Table 12. Summary of Results regarding Organizational Learning Effectiveness.

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

8 Overarching Findings and Thesis Conclusion

Due to advancing digitalization and hardware developments, there are more and more touch points with AI-based services, such as financial robo-advisors or gaming bots, in both private and professional contexts. To enable individuals to take advantage of these AI-based services, they have to accept and use these services. Therefore, the overarching objective of this thesis was to gain a better understanding of individuals' perceptions of AI-based services and how to foster their use as well as the impact of their use on organizations.

Specifically, research papers A and B investigated the differences between the utilization of AI-based services and human-based services; paper C determined the different needs of various user groups; paper D explored methods to promote trust in AI-based services; and paper E analyzed the effect of the utilization of AI-based services on organizational knowledge.

Theoretical Contributions

Thus, all five papers are instrumental in answering the four research questions, providing different theoretical contributions within this thesis.

In response to RQ1, papers A and B show that the advice of AI-based services is perceived and used differently than that of human-based services. This was shown in both high-stakes decisions, such as investment decisions, and low-stakes decisions, such as guessing games. In both examples, AI-based services were followed more than human services, but this is context-specific and cannot be generalized. Thus, other studies in other contexts have yielded opposing results (Castelo et al. 2019; Jussupow et al. 2020). However, the results show that the specific benefits of AI-based services are recognized and accepted by individuals in certain contexts, so such services are used. This could be because AI-based services are not only based on statistical and mathematical calculations, like previous computer programs, but also because the interaction is more natural due to its natural language processing and speech synthesis abilities.

Moreover, by combining the judge-advisor system paradigm and technology acceptance literature, papers A and B extended the IS research through the development of a task-advisor fit that can accurately predict how an advisor is perceived for a specific task. By comparing the fits of the different advisors, a preference for the advisors can be predicted. The task-advisor fit is tested and validated in two different contexts (high-stakes and low-stakes decisions) and it can be seen that the TAF is a predictor for advice utilization. Interestingly, due to the advisors' strengths and weaknesses, different advisor characteristics seem to be crucial for a fit, so the reason for the fits needs to be continually reexamined.

However, paper B also shows that task advisor fit not only has an effect on advice utilization but that this effect is moderated by the characteristics of the advice itself. Hence, the distance of the advice from one's self-assessment influences advice utilization. If the distance is great, advisors have an equally strong influence. This is also reflected in the JAS literature: When judges receive advice that is far from their estimation based on the stimulus-response model, they feel compelled to adjust their evaluations (Schultze et al. 2015). However, some researchers show contrary findings: When it comes to considering advice, suggestions that are close to the initial estimation are more likely to be taken into account than those that are far from it (Minson et al. 2011; Yaniv 2004a), but the results of research paper B support the first case.

Paper C identified factors specific to the acceptance and use of health chatbots for older people. It demonstrated the acceptance and utilization of AI-based services differ depending on the user groups by comparing and extending the results of the study by Laumer et al. (2019) in response to RQ2. Of course, this was only shown for one context and should be tested further in different AI-based services and user groups. Nevertheless, this example shows that the acceptance and use of AI-based systems differ in different user groups. For example, it was shown that older people in particular have a greater demand for emotional support than younger people and that older people may have problems with health chatbots and could not use them if they did not consider these users' medication or medical history. The lack of emotional support provided by machines, which offer solely rational, concise diagnoses, may impede the acceptance and use of health chatbots among older people depending on their health conditions. Future research should investigate how to provide emotional support via machines.

Paper D developed a ranking of mechanisms to increase trust in AI-based advisory services to answer RQ3. There are many varied mechanisms to increase trust, such as transparency (Gönül et al. 2006; Zanker 2012) or enabling testing (Rogers 2003). Organizations generally do not have the resources to implement all of these mechanisms. Paper D demonstrated that the impact of various mechanisms on trust in AI-based services varies and an effective way to promote AI-based services and increase trust is to enable a way to test the services without risk.

Additionally, paper D showed that the judge-advisor system and innovation adoption literature need to be combined to obtain a better understanding of the acceptance and utilization of AI-based services. While the literature on the judge-advisor system regards trust as the most significant factor influencing the adoption of advice (Van Swol and Sniezek 2005), the results show that the relative advantage of informational convivence is more decisive, as described in the innovation adoption research (Choudhury and Karahanna 2008). It is not sufficient to consider the judge-advisor system literature; rather, it should be combined with the innovation adoption literature.

To answer RQ4, paper E simulated an organization's knowledge level to analyze the effect of the use of AI-based services on human and organizational knowledge as well as the effect of needed learning strategies. The results show that using AI-based services leads to a takeover of explorative tasks so that humans within an organization are free to learn more exploitatively or exploratively, whatever they prefer. But the results also show that humans should never be taken completely "out of the loop," even if tasks are largely automated, because human task knowledge is needed for reconfigurations of the AI-based services. Acquiring high levels of organizational knowledge requires at least a moderate amount of reconfiguration effort. Reconfiguration of ML systems cannot be left solely to the IT department as it requires a deep, expert understanding of the domain. Organizations operating in turbulent environments can benefit from relying on knowledge generated by AI-based services, thereby reducing the need for drastic actions such as forced layoffs.

Practical Contributions

This investigation and deep exploration of the acceptance and utilization of AI-based services not only make theoretical contributions to the IS literature but also make practical contributions to organizations with AI-based services and their employees.

In the context of organizations that offer an AI-based service, papers A, B, C, and D provide valuable practical insights and contributions. Papers A and B developed and tested the task-advisor fit, which can be used by organizations to perform a market analysis to predict if their AI-based services will be accepted and used. Thus, it is possible to find out whether there is acceptance of an AI-based service before the expensive development of this service occurs. If such a service is rejected, it is possible to influence various characteristics, such as perceived performance, by providing key performance indicators, which enable a simpler evaluation of the (historical) performance to increase acceptance. However, paper C showed that different user groups have different needs for the acceptance and utilization of AI-based services, so different characteristics of such services may be relevant to their acceptance and utilization for each user group. Various mechanisms can be implemented to influence such characteristics. The deployment of each mechanism incurs significant financial costs and necessitates intensive resources such as human capital. Therefore, it is not possible to implement all of them. Paper D compared different mechanisms to increase trust in AI-based services to increase their effectiveness so that organizations can prioritize their implementation. By prioritizing the implementation of mechanisms that offer the greatest impact on users, organizations can ensure the most impactful and advantageous deployment of resources.

Paper E provided practical insights and contributions for organizations that use AI-based services internally and for their employees. Using AI-based services enables employees to learn more freely. However, their knowledge is still essential because it is needed to reconfigure the AI-based services; human resources are still important for each task and need to be fostered. Employees should not fear that such services will replace them; they will allow them to learn more freely. The positive effect of AI-based services within an organization can be used within a turbulent market, which changes over time.

Concluding Remarks

This thesis analyzed the acceptance and utilization of AI-based services by individuals as well as organizations. Five studies were conducted to answer the four research questions. They showed that AI-based services could be accepted and used and this could be predicted by analyzing their fit with the task. However, the user market should be looked at carefully, as different user groups can have significantly different needs. Not every mechanism that increases acceptance can be implemented and is equally effective, and organizations need to prioritize them. The implementation of such services within an organization often has a positive effect from an organizational point of view, allowing for flexibility in the choice of a learning strategy also in turbulent environments. However, humans should continue to build knowledge of the tasks being performed by AI-based services because this domain knowledge is necessary to reconfigure such services. These studies have their limitations, of course; for example, their contexts are specialized or they have adopted simplifications to emulate AI-based services. Therefore, further studies should be conducted to confirm the results in other contexts and to extend them by accounting for different task characteristics or simulating the interaction of AI-based services.

References

Ada Health. 2019. "Your Personal Health Guide." (https://ada.com, accessed May 28, 2019).

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

Amato, F., Marrone, S., Moscato, V., Piantadosi, G., Picariello, A., and Sansone, C. 2017. "Chatbots Meet Ehealth: Automatizing Healthcare," in *Workshop on Artificial Intelligence with Application in Health*, D. and Impedovo and G. Pirlo (eds.), Bari, Italy.

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

Babylon. 2019. "Mission." (https://www.babylonhealth.com/about, accessed May 28, 2019).

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

Bibault, J.-E., Chaix, B., Nectoux, P., Pienkowski, A., Guillemasé, A., and Brouard, B. 2019. "Healthcare Ex Machina: Are Conversational Agents Ready for Prime Time in Oncology?," *Clinical and Translational Radiation Oncology* (16), The Authors, pp. 55–59. (https://doi.org/10.1016/j.ctro.2019.04.002).

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. (https://doi.org/10.1016/j.obhdp.2006.07.001).

Brandtzaeg, P. B., and Følstad, A. 2017. "Why People Use Chatbots," in *International Conference on Internet Science* (Vol. 10673 LNCS), Springer Verlag, pp. 377–392. (https://doi.org/10.1007/978-3-319-70284-1_30/TABLES/1).

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*.

Bryman, A. 2016. *Social Research Methods*, (5th ed.), Oxford: Oxford University Press.

Brynjolfsson, E., and Mitchell, T. 2017. "What Can Machine Learning Do? Workforce Implications," *Science* (358:6370), pp. 1530–1534. (https://doi.org/10.1126/science.aap8062).

Burrell, J. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms," *Big Data & Society* (3:1), SAGE PublicationsSage UK: London, England, pp. 1–12. (https://doi.org/10.1177/2053951715622512).

Calero Valdez, A., and Ziefle, M. 2019. "The Users' Perspective on the Privacy-Utility Trade-Offs in Health Recommender Systems," *International Journal of Human-Computer Studies* (121), pp. 108–121. (https://doi.org/10.1016/j.ijhcs.2018.04.003).

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), Kluwer Academic Publishers, pp. 7–42. (https://doi.org/10.1023/A:1007850605129).

Carlsson, C., and Walden, P. 2015. "Digital Wellness for Young Elderly: Research Methodology and Technology Adaptation," in *BLED*, Bled, Slovenia, pp. 239–250.

Cassauwers, T. 2020. "Opening the 'Black Box' of Artificial Intelligence," *The EU Research & Innovation Magazine*, , December 1. (https://ec.europa.eu/research-and-innovation/en/horizon-magazine/opening-black-box-artificial-intelligence, accessed November 8, 2022).

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

Cattel, R. B. 1963. "Theory of Fluid and Crystallized Intelligence: A Critical Experiment," *Journal of Educational Psychology* (54:1), pp. 1–22.

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.

Chang, H. H. 2010. "Task-Technology Fit and User Acceptance of Online Auction," *International Journal of Human Computer Studies* (68:1–2), pp. 69–89.

Charlesworth, C. J., Smit, E., Lee, D. S. H., Alramadhan, F., and Odden, M. C. 2015. "Polypharmacy Among Adults Aged 65 Years and Older in the United States: 1988-2010," *Journals of Gerontology: Medical Sciences* (70:8), pp. 989–995.

Chen, K., Hoi, A., and Chan, S. 2014. "Gerontechnology Acceptance by Elderly Hong Kong Chinese: A Senior Technology Acceptance Model (STAM)," *Ergonomics* (57:5), pp. 635–652. (https://doi.org/10.1080/00140139.2014.895855).

Choudhury, P., Allen, R. T., and Endres, M. G. 2021. "Machine Learning for Pattern Discovery in Management Research," *Strategic Management Journal* (42:1), pp. 30–57. (https://doi.org/10.1002/smj.3215).

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

Codish, D., and Ravid, G. 2014. "Personality Based Gamification: How Different Personalities Percive Gamification," in *Proceedings of the European Conference on Information Systems (ECIS)*, Tel Aviv.

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), Management Information Systems Research Center, University of Minnesota, pp. 355–366.

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.

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.

Davison, R. M., and Martinsons, M. G. 2015. “Context Is King! Considering Particularism in Research Design and Reporting,” *Journal of Information Technology* (31:3), pp. 241–249.

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

Dess, G. G., and Beard, D. W. 1984. “Dimensions of Organizational Task Environments,” *Administrative Science Quarterly* (29:1), pp. 52–73.

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.

Dzielinski, M. 2012. “Measuring Economic Uncertainty and Its Impact on the Stock Market,” *Finance Research Letters* (9:3), Elsevier, pp. 167–175. (https://doi.org/10.1016/J.FRL.2011.10.003).

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), NIH Public Access, pp. 98–121. (https://doi.org/10.1016/j.obhdp.2007.05.002).

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*.

Elliott, R., and Timulak, L. 2015. “Descriptive and Interpretive Approaches to Qualitative Research,” in *A Handbook of Research Methods for Clinical and Health Psychology*, J. Miles and P. Gilbert (eds.), Oxford University Press, pp. 147–159. (https://doi.org/10.1093/med:psych/9780198527565.003.0011).

Eule, A. 2017. “Rating the Robo-Advisors,” *Barron’s - The Dow Jones Business and Financial Weekly*.

Eurostat. 2018. “Internet Access and Use Statistics - Households and Individuals.”

Fadhil, A., and Schiavo, G. 2019. “Designing for Health Chatbots,” *ArXiv*.

Feldman, M. S., and Pentland, B. T. 2003. “Reconceptualizing Organizational Routines as a Source of Flexibility and Change,” *Administrative Science Quarterly* (48:1), pp. 94–118. (https://doi.org/10.2307/3556620).

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), Taylor & Francis Group, pp. 67–85.

Fernandes, A. 2017. “Chatbots vs Apps: The Final Frontier,” *Chatbots Magazine*. (https://chatbotsmagazine.com/chatbots-vs-apps-the-final-frontier-a0df10861c48, accessed April 9, 2022).

Flick, U. 2004. “Triangulation in Qualitative Research,” in *A Companion to Qualitative Research*, E. von Kardorff, U. Flick, and I. Steinke (eds.), SAGE Publications, pp. 178–183.

Følstad, A., and Brandtzaeg, P. B. 2020. “Users’ Experiences with Chatbots: Findings from a Questionnaire Study,” *Quality and User Experience* (5:3). (https://doi.org/10.1007/s41233-020-00033-2).

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), American Marketing

Association, 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), American Marketing Association, pp. 39–50.

Gao, J., and Koronios, A. 2010. “Mobile Application Development for Senior Citizens,” in *Proceedings of the 14th Pacific Asia Conference on Information Systems*, Taipei, Taiwan, pp. 214–223.

Gefen, D., Karahanna, E., and Straub, D. W. 2003. “Trust and TAM in Online Shopping: An Integrated Model,” *MIS Quarterly* (27:1), Management Information Systems Research Center, University of Minnesota, p. 51.

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. (https://doi.org/10.1002/bdm.539).

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), Academic Press, pp. 287–302. (https://doi.org/10.1016/J.OBHDP.2008.08.002).

Glass, T. A., and Maddox, G. L. 1992. “The Quality and Quantity of Social Support: Stroke Recovery as Psycho-Social Transition,” *Social Science and Medicine* (34:11), Pergamon, pp. 1249–1261. (https://doi.org/10.1016/0277-9536(92)90317-J).

Glasser, A., and Campbell, M. C. W. 1998. “Presbyopia and the Optical Changes in the Human Crystalline Lens With Age,” *Vision Research* (38:2), Pergamon, pp. 209–229. (https://doi.org/10.1016/S0042-6989(97)00102-8).

Gnewuch, U., Morana, S., and Maedche, A. 2017. “Towards Designing Cooperative and Social Conversational Agents for Customer Service,” in *Proceedings of 38th International Conference on Information Systems*, Seoul, South Korea.

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.

Goodhue, D. L. 1995. “Understanding User Evaluations of Information Systems,” *Management Science* (41:12), INFORMS , pp. 1827–1844. (https://doi.org/10.1287/mnsc.41.12.1827).

Goodhue, D. L., Klein, B. D., and March, S. T. 2000. “User Evaluations of IS as Surrogates for Objective Performance,” *Information & Management* (38:2), North-Holland, pp. 87–101. (https://doi.org/10.1016/S0378-7206(00)00057-4).

Goodhue, D. L., and Thompson, R. L. 1995. “Task-Technology Fit and Individual Performance,” *MIS Quarterly* (19:2), pp. 213–236. (https://doi.org/10.2307/249689).

Grand Review Research. 2018. “MHealth Market Size Worth $151.57 Billion By 2025,” *Grand View Research*. (https://www.grandviewresearch.com/press-release/global-mhealth-market, accessed May 29, 2019).

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

Grönroos, C. 2001. *Service Management and Marketing: A Customer Relationship Management Approach*, (2nd ed.), Wiley, New York, NY.

Guan, C., Wei, J., L Chan, C. M., H Chia, A. B., L, C. M., and H, A. B. 2017. “Senior Citizens' Self-Efficacy for ICT Use: The Influence of Gender, Social Influence and Social Support,” in *Proceedings of The 17th International Conference on Electronic Business*, ICEB, Dubai, UAE, pp. 234–240.

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), Academic Press, pp. 117–133. (https://doi.org/10.1006/OBHD.1997.2697).

Hatch, N. W., and Dyer, J. H. 2004. "Human Capital and Learning as a Source of Sustainable Competitive Advantage," *Strategic Management Journal* (25:12), pp. 1155–1178. (https://doi.org/10.1002/smj.421).

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

Hoermann, S., McCabe, K. L., Milne, D. N., and Calvo, R. A. 2017. "Application of Synchronous Text-Based Dialogue Systems in Mental Health Interventions: Systematic Review," *Journal of Medical Internet Research* (19:8), pp. 1–10.

Hoffmann, L., and Krämer, N. C. 2013. "Investigating the Effects of Physical and Virtual Embodiment in Task-Oriented and Conversational Contexts," *International Journal of Human Computer Studies* (71:7–8), pp. 763–774.

Holgersson, J., and Söderström, E. 2019. "Bridging the Gap - Exploring Elderly Citizens' Perceptions of Digital Exclusion," in *Proceedings of the 27th European Conference on Information Systems*, Stockholm-Uppsala, Sweden.

Hoque, R., and Sorwar, G. 2017. "Understanding Factors Influencing the Adoption of MHealth by the Elderly: An Extension of the UTAUT Model," *International Journal of Medical Informatics* (101), pp. 75–84.

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. (https://doi.org/10.1177/1094670517752459).

Jordan, M. I., and Mitchell, T. M. 2015. "Machine Learning: Trends, Perspectives, and Prospects," *Science* (349:6245), pp. 255–260. (https://doi.org/10.1126/science.aaa8415).

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*, Maui, HI, USA.

Jussupow, E., Benbasat, I., and Heinzl, A. 2020. "Why Are We Averse towards Algorithms? A Comprehensive Literature Review on Algorithm Aversion," in *Twenty-Eighth European Conference on Information Systems*, A Virtual AIS Conference.

Kenny, G., and Connolly, R. 2017. "Towards an Inclusive World: Exploring M-Health Adoption Across Generations," in *Proceedings of the 25th European Conference on Information Systems*, Guimarães, Portugal, pp. 1129–1144.

Kettinger, W. J., and Lee, C. C. 1997. "Pragmatic Perspectives on the Measurement of Information Systems Service Quality," *MIS Quarterly*, pp. 223–240.

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, pp. 1–10.

Kim, M., Park, Jihye, Jeong, D., Park, Jaehyun, and Han, S. 2018. "TRAVEL READER:Free Independent Travel Application for Active Senior by Design Thinking Approach," in *Proceedings of 22nd Pacific Asia Conference on Information Systems*, Japan.

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.

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), Management Information Systems Research Center, University of Minnesota, pp. 941–960.

Kopp, S., Gesellensetter, L., Krämer, N. C., and Wachsmuth, I. 2005. "A Conversational Agent as Museum Guide - Design and Evaluation of a Real-World Application," in *International Workshop on Intelligent Virtual Agents*, T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, and T. Rist (eds.), Berlin, Heidelberg: Springer, pp. 329–343. (https://doi.org/10.1007/11550617_28).

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

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.

Krys, C., and Fuest, K. 2017. "Roland Berger Trend Compendium 2030."

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.

Lally, M. 2016. "Are Women Better Decision Makers than Men? ," *Telegraph.Co.Uk*, , July 3. (https://www.telegraph.co.uk/women/life/are-women-better-decision-makers-than-men/, accessed November 8, 2022).

Laumer, S., Maier, C., and Gubler, F. T. 2019. "Chatbot Acceptenace in Healthcare: Explainin User Adoption of Conversational Agents for Disease Diagnosis," in *Proceedings of 27th European Conference on Information Systems*, Stockholm & Uppsala, Sweden.

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

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), Elsevier, pp. 130–141.

Lee, M. K. 2018. "Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management," *Big Data & Society* (January–Ju), pp. 1–16. (https://doi.org/10.1177/2053951718756684).

Levinthal, D. A., and March, J. G. 1993. "The Myopia of Learning," *Strategic Management Journal* (14:S2), pp. 95–112. (https://doi.org/10.1002/smj.4250141009).

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

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), pp. 0–0.

Lieder, I., Segal, M., Avidan, E., Cohen, A., and Hope, T. 2019. "Learning a Faceted Customer Segmentation for Discovering New Business Opportunities at Intel," in *IEEE International Conference on Big Data*, , December.

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), Pergamon, pp. 252–260.

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

Logg, J. M., Minson, J. A., and Moore, D. A. 2019. "Algorithm Appreciation: People Prefer Algorithmic to Human Judgment," *Organizational Behavior and Human Decision Processes* (151), pp. 90–103. (https://doi.org/10.1016/j.obhdp.2018.12.005).

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. (https://doi.org/10.1016/j.jsis.2016.06.002).

Lu, H.-P., and Yang, Y.-W. 2014. "Toward an Understanding of the Behavioral Intention to Use a Social Networking Site: An Extension of Task-Technology Fit to Social-Technology Fit," *Computers in Human Behavior* (34:43), Pergamon, pp. 323–332.

Maniaci, M. R., and Rogge, R. D. 2014. "Caring about Carelessness: Participant Inattention and Its Effects on Research," *Journal of Research in Personality* (48:1), pp. 61–83. (https://doi.org/10.1016/j.jrp.2013.09.008).

March, J. G. 1991. "Exploration and Exploitation in Organizational Learning," *Organization Science* (2:1), pp. 71–87. (https://doi.org/10.1287/orsc.2.1.71).

Marcinowicz, L., Pawlikowska, T., and Oleszczyk, M. 2014. "What Do Older People Value When They Visit Their General Practitioner? A Qualitative Study," *European Journal of Ageing* (11:4), Springer, pp. 361–367. (https://doi.org/10.1007/s10433-014-0313-0).

Masud, M. M., Gao, J., Khan, L., Han, J., and Thuraisingham, B. 2008. "A Practical Approach to Classify Evolving Data Streams: Training with Limited Amount of Labeled Data," in *Eighth IEEE International Conference on Data Mining*, pp. 929–934. (https://doi.org/10.1109/ICDM.2008.152).

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. (https://doi.org/10.2307/258792).

McCarthy, J. 2007. *What Is Artifcial Intelligence?*

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), Online First Publication, pp. 437–455. (https://doi.org/10.1037/a0028085).

Meisenzahl, M. 2019. "Review: Oral-B Genius X AI Toothbrush," *Businessinsider.Com*, , November 7. (https://www.businessinsider.com/review-oral-b-genius-x-ai-toothbrush-2019-11, accessed November 22, 2022).

Minson, J. A., Liberman, V., and Ross, L. 2011. "Two to Tango," *Personality and Social Psychology Bulletin* (37:10), SAGE PublicationsSage CA: Los Angeles, CA, pp. 1325–1338.

Minutolo, A., Esposito, M., and De Pietroa, G. 2017. "A Conversational Chatbot Based on Kowledge-Graphs for Factoid Medical Questions," *Frontiers in Artificial Intelligence and Applications* (297), pp. 139–152. (https://doi.org/10.3233/978-1-61499-800-6-139).

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

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.

Myers, M. D. 2013. *Qualitative Research in Business and Management*, (2nd ed.), Los Angeles: SAGE Publications.

Nguyen, H. T. T., Tapanainen, T., Theng, Y.-L., Lundberg, S., and Luimula, M. 2015. "Fostering Communication between the Elderly and the Youth with Social Games," (Vol. 227).

Nguyen, T. T. H., Tapanainen, T., and Obi, T. 2014. "A Review of Information and Communication Technology (ICT) Training for Elderly People-Toward Recommendations for Developing Countries," in *Proceedings of the 18th Pacific Asia Conference on Information Systems*, Chengdu, China.

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), Elsevier, 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. (https://doi.org/10.1002/bdm.637).

Panetta, K. 2018. "Gartner Top Strategic Predictions for 2019 and Beyond," *Gartner*. (https://www.gartner.com/smarterwithgartner/gartner-top-strategic-predictions-for-2019-and-beyond/, accessed May 29, 2019).

Parkes, A. 2013. "The Effect of Task–Individual–Technology Fit on User Attitude and Performance: An Experimental Investigation," *Decision Support Systems* (54:2), North-Holland, pp. 997–1009. (https://doi.org/10.1016/J.DSS.2012.10.025).

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), Routledge, 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. (https://doi.org/10.2753/MIS0742-1222290401).

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

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.

Qiu, L., and Benbasat, I. 2009. "Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems," *Journal of Management Information Systems* (25:4), pp. 145–182. (https://doi.org/10.2753/MIS0742-1222250405).

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.

Raisch, S., Birkinshaw, J., Probst, G., and Tushman, M. L. 2009. "Organizational Ambidexterity: Balancing Exploitation and Exploration for Sustained Performance," *Organization Science* (20:4), pp. 685–695. (https://doi.org/10.1287/orsc.1090.0428).

Ransbotham, S., Khodabandeh, S., Kiron, D., Candelon, F., Chu, M., and Lafountain, B. 2020. "Expanding AI's Impact with Organizational Learning," *MIT Sloan Management Review and Boston Consulting Group*, pp. 1–15. (https://sloanreview.mit.edu/projects/expanding-ais-impact-with-organizational-learning/, accessed February 5, 2021).

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

Reynolds, T. J., and Olson, J. C. 2001. *Understanding Consumer Decision Making: The Means-End Approach to Marketing and Advertising Strategy*, Mahwah, NJ, USA: Psychology Press.

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?: Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, United States, pp. 1135–1144. (https://doi.org/10.1145/2939672.2939778).

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. (https://doi.org/10.17705/1CAIS.03026).

Rogers, E. M. 2003. *Diffusion of Innovations*, Free Press.

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), Academy of Management, pp. 393–404. (https://doi.org/10.2307/259285).

Russell, S. J., and Norvig, P. 2016. *Artificial Intelligence: A Modern Approach*, (3rd ed.), Boston, MA, United States: Addison Wesley.

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

Publications.

Samek, W., Wiegand, T., and Müller, K.-R. 2017. “Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models,” *ArXiv:1708.08296*. (https://doi.org/10.48550/arxiv.1708.08296).

Sanyal, S. 2018. “How Is AI Revolutionizing Elderly Care,” *Forbes*. (https://www.forbes.com/sites/shourjyasanyal/2018/10/31/how-is-ai-revolutionizing-elderly-care/#44775f8e07d6, accessed May 29, 2019).

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

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.

Seeger, A.-M., Pfeiffer, J., and Heinzl, A. 2018. “Designing Anthropomorphic Conversational Agents: Development and Empirical Evaluation of a Design Framework,” in *Proceedings of the 39th International Conference on Information Systems*, San Francisco, USA.

Seidel, S., Berente, N., Lindberg, A., Lyytinen, K., and Nickerson, J. V. 2019. “Autonomous Tools and Design: A Triple-Loop Approach to Human-Machine Learning,” *Communications of the ACM* (62:1), pp. 50–57. (https://doi.org/10.1145/3210753).

Sell, A., Walden, C., and Walden, P. 2017. “My Wellness as a Mobile App. Identifying Wellness Types among the Young Elderly,” in *Proceedings of the 50th Hawaii International Conference on System Sciences*.

Shawar, B. A., Atwell, E., and Roberts, A. 2005. “FAQchat as an Information Retrieval System,” in *Proceedings of the 2nd Language and Technology Conference*, Pozná, Poland, pp. 274–278.

Shyu, Y. I. L., Tang, W. R., Tsai, W. C., Liang, J., and Chen, M. C. 2006. “Emotional Support Levels Can Predict Physical Functioning and Health Related Quality of Life Among Elderly Taiwanese With Hip Fractures,” *Osteoporosis International* (17:4), Springer, pp. 501–506. (https://doi.org/10.1007/s00198-005-0020-y).

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

Sironi, P. 2016a. *FinTech Innovation: From Robo-Advisors to Goal Based Investing and Gamification*, Chichester, UK: John Wiley & Sons.

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

Sniezek, J. A., and Buckley, T. 1995. “Cueing and Cognitive Conflict in Judge-Advisor Decision Making,” *Organizational Behavior and Human Decision Processes* (62:2), Academic Press, pp. 159–174. (https://doi.org/10.1006/obhd.1995.1040).

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), Academic Press, pp. 288–307. (https://doi.org/10.1006/obhd.2000.2926).

Solís, W. V., Cedillo, P., Parra, J., Guevara, A., and Ortiz, J. 2017. “Intelligent Pillbox: Evaluating the User Perceptions of Elderly People,” in *Proceedings of the 26th International Conference on Information Systems Development*, Cyprus.

Starmind. 2021. “Swisscom Embeds Starmind in Its Work Smart Strategy to Create a Collaborative, Open-Book Workplace,” *Starmind*, , October 19. (https://www.starmind.ai/customers/swisscom, accessed November 22, 2022).

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

1, 2019).

Steele, R., Lo, A., Secombe, C., and Wong, Y. K. 2009. “Elderly Persons’ Perception and Acceptance of Using Wireless Sensor Networks to Assist Healthcare,” *International Journal of Medical Informatics* (78), pp. 788–801. (https://doi.org/10.1016/j.ijmedinf.2009.08.001).

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. (https://doi.org/10.1016/j.ijforecast.2010.03.002).

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. (https://doi.org/10.1348/014466604X17092).

Szalavitz, M. 2012. “Making Choices: How Your Brain Decides,” *Time.Com*, , September 4. (https://healthland.time.com/2012/09/04/making-choices-how-your-brain-decides/, accessed November 8, 2022).

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

Taylor, K. 2015. “Conntected Health: How Digital Technology Is Transforming Health and Social Care,” London, UK.

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.

Thomas, M., and Powers, J. 2022. “The Future of AI: How Artificial Intelligence Will Change the World,” *Builtin.Com*, , July 18. (https://builtin.com/artificial-intelligence/artificial-intelligence-future, accessed November 8, 2022).

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

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

Tsouna-Hadjis, E., Vemmos, K. N., Zakopoulos, N., and Stamatelopoulos, S. 2000. “First-Stroke Recovery Process: The Role of Family Social Support,” *Archives of Physical Medicine and Rehabilitation* (81:7), W.B. Saunders, pp. 881–887. (https://doi.org/10.1053/apmr.2000.4435).

United Nations. 2017. “World Population Ageing,” New York, USA.

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), Management Information Systems Research Center, University of Minnesota, p. 425.

Venkatesh, V., Thong, J. Y., and Xu, X. 2012. “Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology,” *MIS Quartly* (36:1), pp. 157–178.

Vichitvanichphong, S., Talaei-Khoei, A., and Kerr, D. 2017. “Elderly’s Perception about The Value of Assistive Technologies for Their Daily Living: Impacting Factors and Theoretical Support,” in *Proceedings of the 50th Hawaii International Conference on System Sciences*, pp. 3678–3685.

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), Pergamon, pp. 17–27. (https://doi.org/10.1016/0747-5632(95)00016-X).

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

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. (https://doi.org/10.2753/MIS0742-1222230410).

Wang, W., and Siau, K. 2018. “Trust in Health Chatbots,” in *Proceedings of the 39th International Conference on Information Systems*, San Francisco, USA.

Wanner, J., Heinrich, K., Janiesch, C., and Zschech, P. 2020. “How Much AI Do You Require? Decision Factors for Adopting AI Technology,” in *Forty-First International Conference on Information Systems*, India.

Wansink, B., and Sobal, J. 2007. “Mindless Eating,” *Environment and Behavior* (39:1), Sage PublicationsSage CA: Thousand Oaks, CA, pp. 106–123.

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

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

Weizenbaum, J. 1966. “ELIZA - A Computer Program For the Study of Natural Language Communication Between Man and Machine,” *Communications of the ACM* (9:1), pp. 36–71.

White, T. B. 2005. “Consumer Trust and Advice Acceptance: The Moderating Roles of Benevolence, Expertise, and Negative Emotions,” *Journal of Consumer Psychology* (15:2), Elsevier, pp. 141–148. (https://doi.org/10.1207/s15327663jcp1502_6).

Will, R. P. 1992. “Individual Differences in the Performance and Use of an Expert System,” *International Journal of Man-Machine Studies* (37:2), pp. 173–190.

Wilson, C. 2018. “Is It Love or Loneliness? Exploring the Impact of Everyday Digital Technology Use on the Well-Being of Older Adults,” *Ageing Society* (38:7), pp. 1307–1331. (https://doi.org/10.1017/s0144686x16001537).

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

World Health Organization. 2019. “Risk Factors of Ill Health Among Older People,” World Health Organization. (http://www.euro.who.int/en/health-topics/Life-stages/healthy-ageing/data-and-statistics/risk-factors-of-ill-health-among-older-people, accessed June 16, 2019).

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.

Yaniv, I. 2004a. “Receiving Other People’s Advice: Influence and Benefit,” *Organizational Behavior and Human Decision Processes* (93:1), pp. 1–13. (https://doi.org/10.1016/j.obhdp.2003.08.002).

Yaniv, I. 2004b. “The Benefit of Additional Opinions,” *Current Directions in Psychological Science* (13:2), pp. 75–78. (https://doi.org/10.1111/j.0963-7214.2004.00278.x).

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. (https://doi.org/10.2307/249686).

Your.MD. 2019. “Your.MD - Health Guide and Symptom Checker.” (https://www.your.md/, accessed May 28, 2019).

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*, Dublin, Ireland, pp. 269–272.

Zhang, B., and Li, Y. 2017. "Wearable Medical Devices Acceptance and Expectance of Senior Population in China," in *Proceedings of The 17th International Conference on Electronic Business*, Dubai, UAE, pp. 241–251.

Zigurs, I., and Buckland, B. K. 1998. "A Theory of Task/Technology Fit and Group Support Systems Effectiveness," *MIS Quarterly* (22:3), 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), ACM, pp. 34–50. (https://doi.org/10.1145/344241.344244).

Zimmer, J. C., Henry, R. M., and Butler, B. S. 2007. "Determinants of the Use of Relational and Nonrelational Information Sources," *Journal of Management Information Systems* (24:3), pp. 297–331.

Zuboff, S. 1988. *In the Age of the Smart Machine: The Future of Work and Power*.

Appendix

Appendix 1. Survey Items (Paper A).

Construct	Item		Adapted from…
Task-Advisor Fit	TAF1	The expert's[5] 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.	
	UEX1	I feel very competent in the above explained task.	(Radel et al. 2011)

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

User's Expertise	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.	

Appendix 2. Constructs Items (Paper B).

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.

Appendix 3. Survey items (Paper D).

Item		Adapted from…
TAF1	The expert's[6] 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.	
COM1	The expert is competent and effective in estimating the amount of tic tacs.	(McKnight et al. 2002)
COM2	The expert performs its role of estimation the amount of tic tacs very well.	
COM3	Overall, the expert is a capable and proficient advisor for estimating the amount of tic tacs.	
COM4	In general, the expert is very knowledgeable about the Tic Tacs noise analysis.	
INT1	The expert provides unbiased product recommendations.	(Komiak and Benbasat 2006)
INT2	The expert is honest.	
INT3	I consider the expert to be of integrity.	
EFF	The expert increases the efficiency of my decision making.	(Chan et al. 1997)
BEN1	The expert gives you individual attention.	(Kettinger and Lee 1997)
BEN2	The expert gives you personal attention.	
BEN3	The expert has your best interests at heart.	
BEN4	The expert understands your specific needs.	
EXPL	The advice I get from the expert is easy to comprehend.	(Zimmer et al. 2007)

[6] Depending on the experimental group, the term "expert" is replaced by "human expert" or "AI-based expert" in all items.