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

Neda Mesbah
TU Darmstadt
mesbah@is.tu-darmstadt.de

Christoph Tauchert
TU Darmstadt
tauchert@is.tu-darmstadt.de

Peter Buxmann
TU Darmstadt
buxmann@is.tu-darmstadt.de

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.

1. 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 [1], 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 [2]. 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 [3:2, 4]. AI differs significantly from other technologies, since these AI-based systems have the ability to learn and not just follow static rules [5].

AI-based advisors are often called robo-advisors. They are, compared to humans, only machines that simulate the learning abilities of humans [3] 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 [6, 7]. Often the models used to generate an advice are not interpretable by neither the user nor the developer [8]. 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 [9] 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 [10] 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 [11]. Some studies show that AI-based systems are preferred in contrast to human experts [e.g., 10, 12] and other find contradictory results [e.g., 13, 14].

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 [e.g., 15, 16]. 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

URI: https://hdl.handle.net/10125/71113
978-0-9981331-4-0

[17]. By following the approach of Tauchert and Mesbah [10] 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 [18] for investigations about the user's[1] 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.

2. Advice utilization

People use advice for three main reasons: improvement of their judgement, sharing of responsibility and simply refusal to completely reject received advice [19]. One paradigm that is used in behavioral psychology to investigate advice-taking behavior is the judge-advisor system [20]. 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 [21]. 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 [22].

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 [15, 20, 23, 24, 25]. A summary of JAS studies can be found in [22]. The factors can be categorized in four clusters: characteristics of advisor, characteristics of judge, 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 [26] and age [27] 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" [28:712]. There is not one definition but many different, but by now most researchers agree that it is a multidimensional concept [29, 30]. Mayer et al. [28] 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 [28]. The more competent an advisor is perceived, the more willing a judge is to adjust his estimation [e.g., 24, 31]. **Integrity** is defined as the advisor's honesty and promise keeping [32]. The higher the perceived integrity of the advisor is, the more likely it is that the advice will be used [21]. The same applies to the perceived **benevolence** of an advisor [33], which describes how much an advisor cares about the judge and acts in his interest [32].

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 [34, 35, 36].

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) [15, 16], a more recent study shows that it might be a more complex relation. Schultze et al. [24] 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.

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

3. 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 [17] 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" [17]. 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 [e.g., 37, 38]. Next, we adopt the TTF model in the JAS context and derive our hypotheses.

4. Research model

Until now, the JAS was mainly utilized to investigate the interaction between human decision-makers and human advisors [2, 20, 25]. However, there is one study that investigates differences in the utilization of advice when using a statistical model compared to human advice [39]. 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 [40]. Therefore, another study investigates differences in the utilization of advice of human compared to financial robo-advisors [10]. 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 [11]. Some of the studies show an algorithm aversion, i.e. a preference of human advisors [e.g., 13, 14] while other studies show a preference of AI-based advisors [e.g., 10, 12]. 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. [24] 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 [17]. 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 [17]. 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., 24, 31]. 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 1. After we have derived the research model, the next chapter presents the research method we used to test our model.

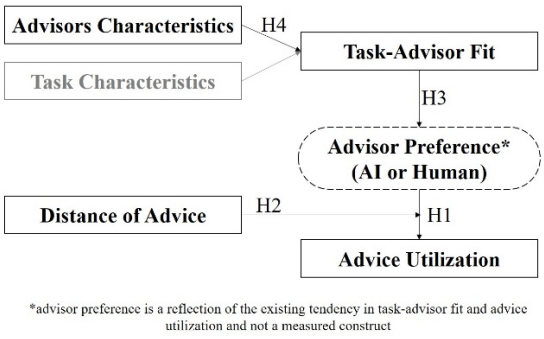

Figure 1. Research model

5. 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 [18] 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 [41]. 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 [25].

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. [42]). 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 [43]. Afterwards, they were introduced to the game.

5.1. 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, Rakotoarisoa and Schulz-Hardt [24], 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.

5.2. 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., 24, 25, 39, 44]:

$$WOA = |final\ estimate - initial\ estimate| \,/\, |advice - initial\ estimate|$$

The weight of advice is a measure that determines to what extent participants consider (weight) an advice in their estimation [15]. 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.

5.3. 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 [45] 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 [29] 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, Choudhury and Kacmar [32] and trust in benevolence of advisor based on a four item scale of Kettinger and Lee [46]. 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. [47]: “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. [48]. All of our construct measurements can be found in Table 4. We also measured tendency towards fantasizing as marker variable to counteract common method bias [43] based on three item scale of Darrat et al. [49].

6. 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 [50, 51]. 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 [52]. 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 [15, 25, 53] 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 1. 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 1), the AI-based advisor is perceived significantly more competent.

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 1. 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.

7. 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 [50]. 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 [52]. 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 [54, 55]. This suits well for theories in their early stages like ours [56].

To asses our measurement model we examined convergent and discriminant validity of the research model [57]. 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 [58]. The item loadings were reported in Table 2. All items have higher loadings than 0.7 as recommended by Hair et al. [57] so that our items are of sufficient

reliability. As can be seen in Table 3, for all constructs Cronbach's α and composite reliability reach the threshold of 0.7 and AVE of 0.5 [59]. 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 [60]. Dess and Beard [61] even set the cut-off value for Cronbach's α to 0.6 for explorative studies.

Table 2. 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

Table 1. 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

Table 3. 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 [57]. To establish discriminant validity, we assessed the cross loadings as well as the square root of the AVE for each construct model [62]. As reported in Table 4 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 4. 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**

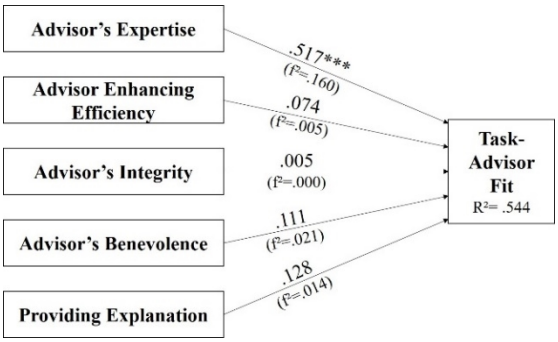

Figure 2. Result of structural model testing
(* p<0.001; ** p<0.01; * p<0.05)**

We depict the results of the research model by running a bootstrapping with 5,000 re-samples [63] in Figure 2. The model fit SRMR is .041, which refers to a good model fit since it is under the cut-off value of .08 [64]. 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.

8. 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 [10] or by Logg et al. [12]. 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 [65], 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 [66]. By transferring this finding into the JAS literature, we show similar to Tauchert and Mesbah [10] 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 [15, 16] However, there is also research with contrary findings, which conclude that more distant advice is given more weight [24]. 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. [24] 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.

9. 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 [18]. Thus, expert systems should be preferred by users with little experience and knowledge for the given task [67]. 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 [25] 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.

10. Appendix

Table 5. Survey items

Item & Adapted from…		
TAF1	The expert's[2] advisory service is compatible with all aspects of this task.	[45]
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.	[32]
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.	[29]
INT2	The expert is honest.	
INT3	I consider the expert to be of integrity.	
EFF	The expert increases the efficiency of my decision making.	[47]
BEN1	The expert gives you individual attention.	[46]
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.	[48]

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

10. References

[1] Wansink, B., and J. Sobal, "Mindless Eating", *Environment and Behavior 39*(1), 2007, pp. 106–123.
[2] Sniezek, J.A., and L.M. Van Swol, "Trust, confidence, and expertise in a judge-advisor system", *Organizational Behavior and Human Decision Processes 84*(2), 2001, pp. 288–307.
[3] McCarthy, J., "What is artifcial intelligence?", 2007.
[4] Elliot, B., and W. Andrews, *A Framework for Applying AI in the Enterprise*, 2017.
[5] Burrell, J., "How the machine 'thinks': Understanding opacity in machine learning algorithms", *Big Data & Society 3*(1), 2016, pp. 205395171562251.
[6] Jung, D., V. Dorner, F. Glaser, and S. Morana, "Robo-Advisory: Digitalization and Automation of Financial Advisory", *Business and Information Systems Engineering 60*(1), 2018, pp. 81–86.
[7] Sironi, P., *The Theory of Innovation: From Robo-Advisors to Goal Based Investing and Gamification*, 2016.
[8] Lipton, Z.C., "The Mythos of Model Interpretability", *ICML Workshop on Human Interpretability in Machine Learning*, (2016).
[9] Simon, H.A., "Theories of Bounded Rationality", *Decision and organization 1*(1), 1972, pp. 161–176.
[10] Tauchert, C., and N. Mesbah, "Following the Robot? Investigating Users' Utilization of Advice from Robo-Advisors", *Fortieth International Conference on Information Systems*, (2019).
[11] Jussupow, E., I. Benbasat, and A. Heinzl, "Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion", *Twenty-Eighth European Conference on Information Systems*, (2020).
[12] Logg, J.M., J.A. Minson, and D.A. Moore, "Algorithm appreciation: People prefer algorithmic to human judgment", *Organizational Behavior and Human Decision Processes 151*, 2019, pp. 90–103.
[13] Castelo, M.W. Bos, and D.R. Lehmann, "Task-dependent algorithm aversion", *Journal of Marketing Research 56*(5), 2019, pp. 809–825.
[14] Lee, M.K., "Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management", *Big Data & Society January–Ju*, 2018, pp. 1–16.
[15] Yaniv, I., "Receiving other people's advice: Influence and benefit", *Organizational Behavior and Human Decision Processes 93*(1), 2004, pp. 1–13.
[16] Minson, J.A., V. Liberman, and L. Ross, "Two to Tango", *Personality and Social Psychology Bulletin 37*(10), 2011, pp. 1325–1338.
[17] Goodhue, D.L., and R.L. Thompson, "Task-Technology Fit and Individual Performance", *MIS Quarterly 19*(2), 1995, pp. 213.
[18] Rzepka, C., and B. Berger, "User Interaction with AI-enabled Systems : A Systematic Review of IS Research", *Thirty Ninth International Conference on Information Systems*, (2018), 1–17.
[19] Harvey, N., and I. Fischer, "Taking Advice: Accepting Help, Improving Judgment, and Sharing Responsibility", *Organizational Behavior and Human Decision Processes 70*(2), 1997, pp. 117–133.
[20] Sniezek, J.A., and T. Buckley, "Cueing and cognitive conflict in judge-advisor decision making", *Organizational Behavior and Human Decision Processes 62*(2), 1995, pp. 159–174.
[21] Van Swol, L.M., "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), 2011, pp. 103–120.
[22] Bonaccio, S., and R.S. Dalal, "Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences", *Organizational Behavior and Human Decision Processes 101*(2), 2006, pp. 127–151.
[23] Van Swol, L.M., and J.A. Sniezek, "Factors affecting the acceptance of expert advice", *British Journal of Social Psychology 44*(3), 2005, pp. 443–461.
[24] Schultze, T., A.-F. Rakotoarisoa, and S. Schulz-Hardt, "Effects of distance between initial estimates and advice on advice utilization", *Judgment and Decision Making 10*(2), 2015, pp. 144–171.
[25] Gino, F., and D.A. Moore, "Effects of task difficulty on use of advice", *Journal of Behavioral Decision Making 20*(1), 2007, pp. 21–35.
[26] Gino, F., J. Shang, and R. Croson, "The impact of information from similar or different advisors on judgment", *Organizational Behavior and Human Decision Processes 108*(2), 2009, pp. 287–302.
[27] Feng, B., and E.L. MacGeorge, "Predicting Receptiveness to Advice: Characteristics of the Problem, the Advice-Giver, and the Recipient", *Southern Communication Journal 71*(1), 2006, pp. 67–85.
[28] Mayer, R.C., J.H. Davis, and F. David Schoorman, "An Integrative Model of Organizational Trust", *The Academy of Management Review 20*(3), 1995, pp. 709–734.
[29] Komiak, S.Y.X., and I. Benbasat, "The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents", *MIS Quarterly 30*(4), 2006, pp. 941–960.
[30] Rousseau, D.M., S.B. Sitkin, R.S. Burt, and C. Camerer, "Not so Different after All: A Cross-Discipline View of Trust", *The Academy of Management Review 23*(3), 1998, pp. 393–404.
[31] Kim, H., I. Benbasat, and H. Cavusoglu, "Online Consumers' Attribution of Inconsistency Between Advice Sources", *Thirty Eighth International Conference on Information Systems*, (2017), 1–10.
[32] McKnight, D.H., V. Choudhury, and C. Kacmar, "Developing and Validating Trust Measures for e-Commerce : An Integrative Typology", *Information Systems Research 13*(3), 2002, pp. 334–359.
[33] White, T.B., "Consumer trust and advice acceptance: The moderating roles of benevolence, expertise, and negative emotions", *Journal of Consumer Psychology 15*(2), 2005, pp. 141–148.
[34] Ye, L.R., and P.E. Johnson, "The Impact of Explanation Facilities on User Acceptance of Expert Systems Advice", *MIS Quarterly 19*(2), 1995, pp. 157.
[35] Wang, W., and I. Benbasat, "Recommendation Agents for Electronic Commerce: Effects of Explanation Facilities on Trusting Beliefs", *Journal of Management Information*

Systems, (2007), 217–246.
[36] Zanker, M., "The influence of knowledgeable explanations on users' perception of a recommender system", *Proceedings of the sixth ACM conference on Recommender systems*, (2012), 269–272.
[37] Robles-Flores, J.A., and D. Roussinov, "Examining Question-Answering Technology from the Task Technology Fit Perspective", *Communications of the Association for Information Systems 30*, 2012, pp. 439–454.
[38] Zigurs, I., and B.K. Buckland, "A Theory of Task/Technology Fit and Group Support Systems Effectiveness", *MIS Quarterly 22*(3), 1998, pp. 313–334.
[39] Önkal, D., P. Goodwin, M. Thomson, S. Gönül, and A. Pollock, "The relative influence of advice from human experts and statistical methods", *Journal of Behavioral Decision Making 22*(4), 2009, pp. 390–409.
[40] Anthes, G., "Artificial intelligence poised to ride a new wave", *Communications of the ACM 60*(7), 2017, pp. 19–21.
[41] Camerer, C.F., and R.M. Hogarth, "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework", *Journal of Risk and Uncertainty 19*(1/3), 1999, pp. 7–42.
[42] Lowry, P.B., J. D'Arcy, B. Hammer, and G.D. Moody, "'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), 2016, pp. 232–240.
[43] Podsakoff, P., S. MacKenzie, J.Y. Lee, and N.P. Podsakoff, "Common method biases in behavioral research: a critical review of the literature and recommended remedies", *J. Appl. Psychol. 88*(5), 2003, pp. 879.
[44] Sah, S., D.A. Moore, and R.J. MacCoun, "Cheap talk and credibility: The consequences of confidence and accuracy on advisor credibility and persuasiveness", *Organizational Behavior and Human Decision Processes 121*(2), 2013, pp. 246–255.
[45] Moore, G.C., and I. Benbasat, "Development of Instrument to Measure the Perceptions of Adopting an Information Technology Innovation", *Information Systems Research 2*(3), 1991, pp. 192–222.
[46] Kettinger, W.J., and C.C. Lee, "Pragmatic Perspectives on the Measurement of Information Systems Service Quality", *MIS Quarterly*, 1997, pp. 223–240.
[47] Chan, Y.E., S.L. Huff, D.W. Barclay, and D.G. Copeland, "Business Strategic Orientation, Information Systems Strategic Orientation, and Strategic Alignment", *Information Systems Research 8*(2), 1997, pp. 125–150.
[48] Zimmer, J.C., R.M. Henry, and B.S. Butler, "Determinants of the Use of Relational and Nonrelational Information Sources", *Journal of Management Information Systems 24*(3), 2007, pp. 297–331.
[49] Darrat, A.A., M.A. Darrat, and D. Amyx, "How impulse buying influences compulsive buying: The central role of consumer anxiety and escapism", *Journal of Retailing and Consumer Services 31*, 2016, pp. 103–108.
[50] Meade, A.W., and S.B. Craig, "Identifying careless responses in survey data", *Psychological Methods 17*(3), 2012, pp. 437–455.
[51] Maniaci, M.R., and R.D. Rogge, "Caring about carelessness: Participant inattention and its effects on research", *Journal of Research in Personality 48*(1), 2014, pp. 61–83.
[52] Eurostat, "Internet Access and Use Statistics - Households and Individuals", 2018.
[53] Yaniv, I., "The Benefit of Additional Opinions", *Current Directions in Psychological Science 13*(2), 2004, pp. 75–78.
[54] Qureshi, I., and D. Compeau, "Assessing Between-Group Differences in Information Systems Research: A Comparison of Covariance- and Component-Based SEM", *MIS Quarterly 33*(1), 2009, pp. 197–214.
[55] Ringle, C.M., S. Wende, and J.-M. Becker, *SmartPLS 3*, SmartPLS GmbH, Hamburg, 2015.
[56] Fornell, C., and F.L. Bookstein, "Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory", *Journal of Marketing Research 19*(4), 1982, pp. 440–452.
[57] Hair, J.J.F., G.T.M. Hult, C. Ringle, and M. Sarstedt, *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Sage Publications, 2013.
[58] Xu, H., T. Hock-Hai, B.C.Y. Tan, and R. Agarwal, "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), 2012, pp. 1342–1363.
[59] Hair, J.F., C.M. Ringle, and M. Sarstedt, "PLS-SEM: Indeed a Silver Bullet", *Journal of Marketing Theory and Practice 19*(2), 2011, pp. 139–152.
[60] Codish, D., and G. Ravid, "Personality based Gamification: How different Personalities percive Gamification", *Proceedings of the European Conference on Information Systems (ECIS)*, (2014).
[61] Dess, G.G., and D.W. Beard, "Dimensions of Organizational Task Environments", *Administrative Science Quarterly 29*(1), 1984, pp. 52–73.
[62] Fornell, C., and D.F. Larcker, "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error", *Journal of Marketing Research 18*(1), 1981, pp. 39–50.
[63] Davison, A.C., and D. V. Hinkley, *Bootstrap Methods and Their Application*, Cambridge University Press, New York, NY, 1997.
[64] Hu, L., and P.M. Bentler, "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives", *Structural Equation Modeling: A Multidisciplinary Journal 6*(1), 1999, pp. 1–55.
[65] Hoffmann, L., and N.C. Krämer, "Investigating the effects of physical and virtual embodiment in task-oriented and conversational contexts", *International Journal of Human Computer Studies 71*(7–8), 2013, pp. 763–774.
[66] Chang, H.H., "Task-technology fit and user acceptance of online auction", *International Journal of Human Computer Studies 68*(1–2), 2010, pp. 69–89.
[67] Will, R.P., "Individual differences in the performance and use of an expert system", *International Journal of Man-Machine Studies 37*(2), 1992, pp. 173–190.