First impressions of a financial AI assistant: differences between high trust and low trust users

Simon Schreibelmayr, Laura Moradbakhti and Martina Mara*

Robopsychology Lab, Linz Institute of Technology, Johannes Kepler University Linz, Linz, Austria

Calibrating appropriate trust of non-expert users in artificial intelligence (AI) systems is a challenging yet crucial task. To align subjective levels of trust with the objective trustworthiness of a system, users need information about its strengths and weaknesses. The specific explanations that help individuals avoid over- or under-trust may vary depending on their initial perceptions of the system. In an online study, 127 participants watched a video of a financial AI assistant with varying degrees of decision agency. They generated 358 spontaneous text descriptions of the system and completed standard questionnaires from the Trust in Automation and Technology Acceptance literature (including perceived system competence, understandability, human-likeness, uncanniness, intention of developers, intention to use, and trust). Comparisons between a high trust and a low trust user group revealed significant differences in both open-ended and closed-ended answers. While high trust users characterized the AI assistant as more useful, competent, understandable, and humanlike, low trust users highlighted the system's uncanniness and potential dangers. Manipulating the AI assistant's agency had no influence on trust or intention to use. These findings are relevant for effective communication about AI and trust calibration of users who differ in their initial levels of trust.

KEYWORDS

human-AI interaction, banking, user perception, trust calibration, acceptance, agency, survey

1. Introduction

When someone else makes financial decisions for you or suggests how to manage your bank account, it is a highly trust-relevant situation. Financial decisions involve risks and may have long-term implications for the account holder. Therefore, it is important to evaluate how trustworthy those who make or propose such decisions are. If one places too much trust in a financial advisor who actually does not deserve the level of trust, for example due to bad intentions or a lack of competence, this can lead to over-reliance, wrong decisions and associated harms for the trustor (Miller et al., 2016; Robinette et al., 2016; Chong et al., 2022). If, on the other hand, one distrusts a financial advisor despite their good intentions and high level of competence, one may miss out on possible benefits or have to allocate alternative resources to reach comparable results (Grounds and Ensing, 2000; Kamaraj and Lee, 2022).

Today, trust-relevant decisions are increasingly made in conjunction with Artificial Intelligence (AI). AI advisors and other algorithmic decision support systems are used in the field of banking and investments (Day et al., 2018; Shanmuganathan, 2020), but also in many other high-risk domains such as medical diagnoses (Dilsizian and Siegel, 2014; Das et al., 2018), human resources (Vrontis et al., 2022; Chowdhury et al., 2023),

mushroom identification (Leichtmann et al., 2023a,b) and even court decisions (Hayashi and Wakabayashi, 2017; Rosili et al., 2021). How to avoid *over-trust*, i.e., unjustified excessive trust in AI systems, and *under-trust*, i.e., insufficient trust in reasonably fair and accurate AI systems, has therefore become a major topic of social scientific, psychological, and ethical research of human-AI collaboration in recent years (Miller et al., 2016; Eigenstetter, 2020). A responsible perspective on human-AI relations can never aim at simply boosting user trust, but rather supporting the calibration of appropriate levels of trust in AI. In this context, *calibrated trust* has been described as trust that is proportionate to the actual trustworthiness (objective capabilities and limitations) of an AI system, rather than trusting excessively or insufficiently (Lee and See, 2004; Boyce et al., 2015; de Visser et al., 2020). Trust calibration involves accurate assessments of an AI system's strengths and weaknesses and adjusting one's subjective level of trust accordingly. Exploring methods for dampening (reducing over-trust) or repairing (reducing under-trust) mis-calibrated trust in AI are therefore of great interest to the scientific community (Lee and See, 2004; de Visser et al., 2020; Chong et al., 2022).

Previous research has found that providing information can contribute to more adequate calibration of trust in AI. By offering more detailed explanations about the strengths and limitations of an AI system, and thus clarifying potential misconceptions and false expectations, users can adjust how much they want to trust the system's output (Boyce et al., 2015; Körber et al., 2018; Miller, 2019; Tomsett et al., 2020; Leichtmann et al., 2023b). At the same time, empirical studies from the field of human-robot interaction indicate that users often have very different initial perceptions and mental models of the same machine and therefore trust or distrust them for very different reasons (Olson et al., 2009; Chien et al., 2016; Matthews et al., 2020). Over the years, hundreds of different cognitive architectures have been described and some of them have been integrated into AI-based models to investigate different approaches, encompassing aspects such as thinking similar to humans, rational thinking, human-like behavior, and rational behavior (Kotseruba et al., 2016; Lieto et al., 2018). While for some users utilitarian dimensions such as the competence, performance, or practicality of the AI system for a specific task might lead to high trust, other users might base their distrust more on social-emotional components such as a low human-likeness or a perceived lack of adherence to social norms. It is likely that the information that could help these users to establish adequate trust would be different.

Consequently, to estimate what kind of explanations about an AI application can effectively contribute to appropriate trust calibration among different user groups, it is important to look at variations in first impressions of the system. The present work therefore investigates how initial perceptions of a financial AI assistant differ between people who report they (rather) trust and people who report they (rather) do not trust the system. In order to capture potentially different user perceptions in a broad and less biased way, we combine typical closed self-report scales with an analysis of qualitative textual descriptions generated by the study participants themselves. We decided to conduct the study in the context of AI in banking, as this is a sensitive domain in terms of risk and vulnerability, making trust a particularly relevant concept (Hannibal, 2023).

1.1. AI as financial advisor

AI assistants have established a place in our lives and are applied millions of times worldwide (Statista, 2021), for which a specific and increasing number is taking place in the financial sector and the banking industry (Kochhar et al., 2019). Artificial Intelligence as an advisor in financial decisions is becoming a reality in everyday business (Ludden et al., 2015; Vincent et al., 2015; Jung et al., 2018; Shanmuganathan, 2020). AI can advise on whether to make bank transfers, deposit money, buy stocks, or to make other investments, sometimes in a highly speculative manner (Fein, 2015; Tertilt and Scholz, 2018). For instance, to promptly respond to price fluctuations, many individuals execute such financial decisions directly on their smartphones or tablets (Varshney, 2016; Azhikodan et al., 2019; Lele et al., 2020). Privacy considerations are crucial (Manikonda et al., 2018; Burbach et al., 2019), and the level of autonomy of a financial bot may impact its acceptance and adoption. As financial AI assistants have access to sensitive data and may be able to make critical decisions independently, users may be hesitant to use it if they perceive a lack of control over the system's actions.

1.2. Trust in AI

Research on human-machine interaction shows that individuals engage in social behavior toward machines by applying heuristics from interpersonal relations (Nass et al., 1997; Christoforakos et al., 2021). Trust is a critical component in social interactions, as it allows individuals to rely on and believe in the actions and statements of others (Rotter, 1980; Scheuerer-Englisch and Zimmermann, 1997; Neser, 2016). Social psychologists view trust as a multidimensional construct that is composed of cognitive, affective, and behavioral dimensions (Johnson and Grayson, 2005; Righetti and Finkenauer, 2011; Lyon et al., 2015). This means that trust involves positive expectations toward a trustee based on an emotional bond, and it expresses itself in concrete actions of both the trustor and the trustee. Trust also depends on the existence of risk and the individual's willingness to accept vulnerability, and it is associated with expectations of a positive outcome and positive prospective behavior from the other person (Mayer et al., 1995; Flores and Solomon, 1998).

Cognitive trust (as one route of trust formation) is characterized by the assessment of an individual's reliability and dependability (Rempel et al., 1985; McAllister, 1995; Colwell and Hogarth-Scott, 2004), based on the trustor's rational evaluation of the trustee's knowledge, competence, or understandability (McAllister, 1995; Kohn et al., 2021). In contrast, affective trust is an advancement of cognitive trust (Chen et al., 1998; Kim, 2005), which involves mutual interpersonal care and concern, and is characterized by an emotional bond and feelings of security between individuals (Rempel et al., 1985; Johnson and Grayson, 2005). Emotional dependence is a critical determinant in which the impact on trust is more closely associated with personal experiences as compared to cognitive trust. Both the emotional aspect and task-specific abilities may lead to either a positive or negative evaluation of an agent's trustworthiness. Components

such as competence and knowledge gained from interpersonal interaction research (McAllister, 1995) are highly important in HCI and affect the level of trust (Chen and Terrence, 2009; Christoforakos et al., 2021) and ultimately the acceptance of automated systems (Ward et al., 2017).

When interacting with digital assistants and AI applications, it may be difficult to accurately evaluate trust-relevant characteristics (Rai, 2020). Often labeled as "black box" services, users have less knowledge about what actually is transpired in the service performance which then requires a high portion of trust (Rosenberg, 1983; MacKenzie, 2005; Devlin et al., 2015). Appropriate trust calibration is one important factor which should be considered in the use of AI powered assistant tools. Low levels of trust can lead to disuse (Lee and Moray, 1992; Parasuraman and Riley, 1997), while very high levels of trust in automated systems may be associated with over-reliance and over-trust (Freedy et al., 2007; Parasuraman et al., 2008; Körber et al., 2018). Additionally, users often evaluate their own decisions in comparison to automated solutions but may also develop a tendency to over-rely on automated systems (Young and Stanton, 2007; Chen and Terrence, 2009). Conversely, people who overestimate their performance tend to exhibit under-reliance on AI systems (similar to the Dunning-Kruger Effect in social psychology), which hinders effective interaction with AI-powered assistant tools (He et al., 2023).

1.3. AI agency

In addition to certain voice characteristics such as gender, naturalness or accent (Nass and Brave, 2005; McGinn and Torre, 2019; Moradbakhti et al., 2022; Schreibelmayr and Mara, 2022), other features on the behavioral level of assistance tools like agency are becoming more and more important (Kang and Lou, 2022). Agency pertains to an AI assistant's self-control ability (e.g., Gray et al., 2007), whereas autonomous AI usually has high agency, making decisions based on data, while low agency AI relies on consistent human input. Machines and systems work on a highly autonomous level and according to the situation, offering their help adaptively and in real time (Qiu et al., 2013; Profactor, 2021). Even though users want to experience a proactive style when interacting with a chatbot (Medhi Thies et al., 2017; Pizzi et al., 2021), they also crave some sense of control over the chatbot's or autonomous system's actions and may feel threatened if they behave too autonomously (Złotowski et al., 2017; Stein et al., 2019; Seeber et al., 2020a,b). By way of example, proactivity for irrelevant information is viewed negatively (Chaves and Gerosa, 2021), but it is largely unclear which degree of agency is actually desirable and demanded for AI. Various studies indicate the effects of agency on interactions between humans and virtual characters or robots (Guadagno et al., 2007; Beer et al., 2014; Fox et al., 2015; Pitardi et al., 2021) and Stein et al. (2019) have shown, for example, that participants experienced significantly stronger eeriness if they perceived an empathic character to be an autonomous AI. Considering the phenomenon of the uncanny valley (MacDorman and Ishiguro, 2006; Mori et al., 2012), the concept of a thinking robot that autonomously generates ideas,

desires, and expresses needs (Parviainen and Coeckelbergh, 2021; Hanson Robotics, 2023) is unsettling and evokes strong feelings of eeriness alongside fascination (Gray and Wegner, 2012; Stein and Ohler, 2017; Appel et al., 2020). The degree of agency differs depending on technical possibilities and improvements and more research has to be done taking various perspectives into account to create a user-centric point of view. Based on the literature, the degree of agency of an AI banking assistant should be considered in the present study to examine the impact on trust and intention to use.

1.4. The present study

In the current research, we asked participants about their perceptions of an artificially intelligent banking assistant. As the utilization of assistance systems in the financial sector is becoming more important and poses potential risks related to trust theories (Mayer et al., 1995; Lee and See, 2004), we chose this particular context to examine participants' impressions after they viewed a video introduction of a fictitious AI banking assistant. We explored impressions that were evoked by the AI banking assistant across different user groups (low vs. high trust) by asking the participants to write down how they would describe the presented AI. Additionally, participants answered closed self-report questionnaires typically used in empirical studies on trust in automation, technology acceptance, and the uncanny valley phenomenon. As experimental manipulation, we incorporated two levels of autonomy into the AI banking assistant. Factors of the technology-acceptance model (TAM, Venkatesh and Bala, 2008) were included and discussed in the light of the Trust in Automation literature (Körber, 2019).

2. Methods

The study described in the following was conducted as part of a larger project with multiple research questions. To ensure transparency, we would like to mention that the present study builds upon a preliminary investigation in which the manipulation check failed. Therefore, we strengthened the manipulation by placing additional emphasis on the textual content in the videos (stimulus) and dividing the videos into two parts to maintain the impression created immediately after the manipulation. Below, we give the characteristics of our sample, the study procedure, and the measures (e.g., trust, intention to use) used. For the sake of completeness, other variables that were also surveyed but are not relevant to the present paper are briefly listed: Voice-realism, pleasantness, perception of risk, tolerance of ambiguity, and desirability for control.

2.1. Sample size justification and participants

The sample size required for the present online experiment was calculated by a power analysis, which helps to determine how many study participants are needed, using G*Power (Cohen, 1992;

Faul et al., 2007). For the calculation, a medium effect size of $d = 0.50$ was assumed and α error probability was set to 0.05. In order to achieve a power (1 – β) of 80% (β represents the probability of committing a Type II error, which is the error of failing to reject a null hypothesis that is actually false), the analysis resulted in a recommended sample size of at least $N = 128$ to run a test procedure with two independent groups. The participants were recruited through a snowball approach[1] at the campus of the Johannes Kepler University Linz, Austria. Overall, 136 participants took part in the online experiment. After excluding 9 participants as their indicated age was below 16 years or because of insufficient German skills, the final sample consisted of 127 participants (62 females, 61 male, 2 non-binary, 2 no specification), aged between 17 and 82 years ($M = 30.55$, SD $= 15.47$). 49.6% of participants had completed compulsory education with a high school diploma (Matura or Abitur), 38.6% had completed university education with a bachelor's level or higher, and 11.8% had some other type of vocational training.

2.2. Study design and procedure

The experiment was conducted as a 2 x 1 between-subjects design, involving 2 groups of participants experiencing distinct levels of the independent variable to analyze its effects on the dependent variable ($N_{Highagency} = 61$, $N_{Lowagency} = 66$), and was designed with the online survey software (Questback, 2020). At the beginning of the study, all participants were instructed to either use headphones or keep their computer/laptop audio on a high volume for the duration of the study. After that, participants read an introduction, confirmed their consent, filled out demographic information (including age, gender, and level of education) and were asked to fill in personality questionnaires (openness to experience, neuroticism, and propensity to trust). Once these scales were completed by the participants, brief instructions appeared followed by one out of two AI banking assistant videos, which were randomly assigned to each participant. Additionally, every video was divided into two parts (∼1 min each). Subsequent to the first part, the participants were queried about the degree of human-likeness and eeriness of the AI banking assistant. Items regarding AI banking assistant's agency (manipulation-check) were asked after the first part of the video as well. After the second part of the video, the participants were requested to assess the perceived level of trustworthiness they had in the assistant featured in the video. Then, the participants were requested to write down attributes of the AI banking assistant and their intention to use it. Finally, some check items were asked to make sure that the participants had clearly understood the sound of the video (it was technically not possible to skip the video), that their German language skills allowed them to understand the content of the study, and that they had answered all questions honestly and conscientiously. The

---

1 Individuals who had already participated were asked to invite new study participants. Persons who had already participated in the online-experiment were sensitized to not communicate any additional information about the contents of the study to newly recruited persons.

average processing time for the entire survey was approximately 15 min and 31 s.

2.3. Stimulus material and manipulations

2.3.1. AI banking assistant videos

We used two videos, differing in the spoken content (low/high agency). The videos were created using the software Adobe After Effects. In order to make the respective condition (low/high agency) stand out, certain content, words and phrases of the spoken text were visualized and highlighted in the video (e.g., "at your request" = low agency, vs. "without your intervention" = high agency). The voice for the AI banking assistant was created with the help of the text-to-speech online platform ttsmp3.com, powered by AWS Polly (Amazon Web Services), based on the default settings of the voice "Vicki". A sound wave was shown, to visualize the AI banking assistant's voice in the video (see Figure 1). The total length of the video was 1 min and 53 s in the high agency condition and 1 min and 52 s in the low agency condition (each video was divided into two parts as we asked some scales in between).

2.3.2. AI banking assistant agency (manipulations)

In the context of our study, "high agency" refers to AI systems that can plan ahead and make decisions without user command, while "low agency" AI relies on consistent human input for tasks and choices. High agency AI operates independently, low agency AI requires ongoing human guidance. We varied the AI banking assistant's agency level between high and low agency through manipulating its textual introduction. In the low agency condition, for example, the AI banking assistant offered: "If you connect me with your account information, I can support you with deposits, planned savings and other tasks." In the high agency condition, the same service was introduced in the following manner: "I will autonomously connect myself with your account information. Without effort on your part, I will take care of deposits, planned savings and other tasks." Overall, 6 phrases were adapted and differed between the conditions to achieve the agency level manipulation.

2.4. Measures

Open-ended descriptions of the AI banking assistant. We asked the participants to provide a verbal description of the AI banking assistant using three terms in an open comment format. ("If you were to tell someone else about the AI banking assistant in the video, what terms would you use to describe it?").

Trust was measured with two items of the trust in automation subscale (Körber, 2019). The items were modified slightly to match the content of the AI banking assistant (e.g., "I would trust the AI banking assistant.", ranging from 1 = *not at all* to 5 = *very much*) (Cronbach's α = 0.760). Find all items in Appendix B.

Competence (4 items, Cronbach's α = 0.598), *understandability* (2 items, Cronbach's α = 0.744), and *intention of developers* (2 items, Cronbach's α = 0.738) were measured with the help of

FIGURE 1
Screenshot from an AI banking assistant video including a text passage (English translation: "I am your artificially intelligent financial assistant") and sound waves.

shortened scales taken from the trust in automation model (Körber, 2019). The items were modified slightly to match the content of the AI banking assistant. Find all items in Appendix B.

Intention to use was measured with the help of two items based on the intention to use items from the Technology Acceptance Model (TAM3, Venkatesh and Bala, 2008). A 5-point Likert-scale (1 = *not at all* to 5 = *very much*) was used and the items were slightly adapted to fit the context of the current experiment: "I could imagine using the AI banking assistant in the future." and "I would like to be informed about products that are similar to the AI banking assistant.". The reliability of the items was high, with a Cronbach's α of 0.804.

Human-likeness of the AI banking assistant, which refers to the extent to which a system resembles human attributes, behavior, appearance, or cognitive capabilities, was assessed with five items on a five-point semantic differential scale [e.g., 1 = *synthetic*, 5 = *real*; 1 = *mechanical*, 5 = *organic*, adapted from Ho and MacDorman (2010), which yielded an excellent reliability with Cronbach's α = 0.832.]

Uncanniness of the AI banking assistant, which describes an unsettling sensation or a sense of discomfort and creepiness (e.g., Mori, 1970; Mori et al., 2012; due to the blurring of human and non-human traits referring to the uncanny hypothesis), was measured with three items on a five-point semantic differential scale [e.g., 1 = *scary*, 5 = *comforting*, as example of an inverse coded item, adapted from Ho and MacDorman (2010), Cronbach's α = 0.837].

Agency, consisting of two items, was used as a manipulation check for the different agency levels (low/high) of the two conditions (e.g., "The AI banking assistant has the ability to act in a self-controlled manner"; Cronbach's α = 0.725). A 5-point Likert-scale was used.

We used the two dimensions of the Big Five personality traits, *openness to experience,* which refers to an individual's inclination to engage with novel ideas, experiences, and intellectual or artistic pursuits (Neyer and Asendorpf, 2018) and *neuroticism,* which is characterized by a tendency toward experiencing negative emotions such as anxiety, depression, moodiness, and emotional instability (Neyer and Asendorpf, 2018). Both scales were explained by three items each, based on the 15-items

short-scale (5 dimensions) from the Socio-Economic Panel (SOEP, see Schupp and Gerlitz, 2014). A 5-point Likert-scale was used. Internal consistencies were good (openness to experience: Cronbach's α = 0.708, neuroticism: Cronbach's α = 0.713).

Propensity to trust, as a personality trait characterized by an individual's innate inclination to believe in the reliability and good intentions of others in various situations (Patent and Searle, 2019), was assessed with the help of three items (5-point Likert-scale) taken from the trust in automation model (Körber, 2019), showing acceptable internal consistency (Cronbach's α = 0.503).

3. Results

Before analysis, we examined if the prerequisites of parametric analyses (normal distribution, homoscedasticity of the variances) were met by our data. As this was not the case for several variables, we decided to apply non-parametric test procedures. Therefore, we used the Wilcoxon-Mann-Whitney-Test to compare the two independent groups and determined if their distributions significantly differs, to assess whether one group tends to have higher or lower values than the other based on the ranking of observations in the samples. Zero-order correlations (Spearman's rank) between all variables were computed and can be found in Appendix Table 1.

Our analysis aimed to compare individuals who exhibit low levels of trust in the AI banking assistant with those who have high levels of trust in it. To accomplish this, we removed participants who provided an average value of 3 on the 5-point Likert-scale for trust in the further evaluation and formed the two extreme groups. All participants with trust scores ranging from 1 to 2.5 were placed in the low trust group, while those with trust scores ranging from 3.5 to 5 were placed in the high trust group (1.00 = 10 VPN, 1.50 = 8 VPN, 2.00 = 26 VPN, 2.50 = 18 VPN, 29 VPN with a score of 3.00 were excluded, 3.50 = 20 VPN, 4.00 = 11 VPN, 4.50 = 3 VPN, 5.00 = 2 VPN). Figure 2 provides a descriptive representation of the two groups, low trust and high trust. The corresponding
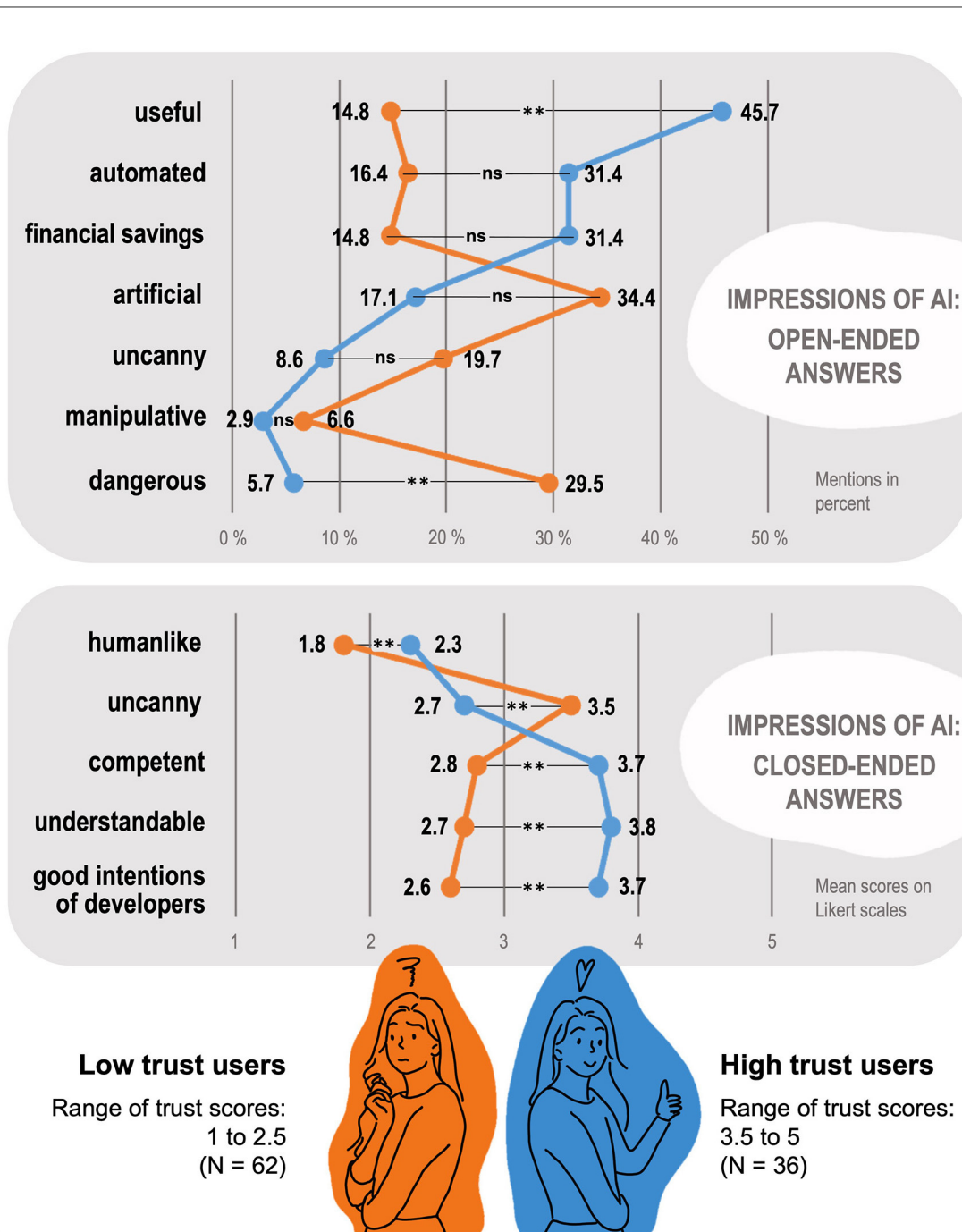
FIGURE 2
Overview of characteristics attributed to the financial AI assistant by users with low trust and users with high trust in the system. Open-ended descriptions represent spontaneous characterizations that users generated themselves, indicated in percentage frequency per group. Closed-ended descriptions represent answers to preset questionnaires, indicated in mean scores per group. Statistical significance of group differences: **$p <$ 0.001; ns = non-significant.

statistical values (e.g., standard errors of means) can be found in Appendix Table 2.

In further analysis, we compared open-ended descriptions of the AI banking assistant (user-generated impressions) and closed-ended scale means (e.g., intention to use) between these two extreme groups. Additionally, we investigated whether there were any differences in the measured personality dimensions between the two user groups. Finally, we tested whether there were any differences in the variables between the two different agency levels (low vs. high) of the AI banking assistants as well. Since the degree of agency (high/low) was manipulated, we also examined the distribution of the two conditions within the two groups low trust vs. high trust. The descriptive analysis indicates that the two agency variants are almost equally distributed within

the two groups low trust and high trust, $X^2(1,98) = 0.70$, $p = 0.791$.

3.1. Analysis of open-ended descriptions

After the participants wrote down three terms (e.g., nouns, adjectives) that described the AI banking assistant in an open comment field, two to three independent raters evaluated these terms. In total, 358 (23 missing) descriptions were recorded. The raters had to determine for each participant whether at least one of the three terms reflected specific categories (useful, automated, financial savings, artificial, uncanny, manipulative, dangerous).

These seven categories were created collaboratively *post-hoc* based on the most frequently occurring categories identified. Per category, each user-generated answer was assigned either 0 (if none of the terms matched the category) or 1 (if at least one of the terms matched the category). In cases where the raters did not agree, a decision was made collaboratively. The interrater reliabilities demonstrated substantial to nearly perfect agreement (Landis and Koch, 1977): Useful: κ = 0.727, automated: κ = 0.977, financial savings: κ = 0.832, artificial: κ = 0.895, uncanny: κ = 0.899, manipulative: κ = 0.725, dangerous: κ = 0.634.

Furthermore, the raters assigned a separate category for an overall impression (sentiment) to each test participant. They rated this category based on whether the terms used were only neutral, at least one positive, at least one negative, or a combination of both negative and positive terms. The sentiment was rated on a scale from −1 (negative) to 0 (neutral) to 1 (positive), depending on the terms used (14 ambivalent combinations including both negative and positive terms were excluded from the analysis, 5 VPN were additionally excluded due the fact that they didn't write down any descriptions). Subsequently, a mean score was calculated based on these evaluations, which represented the overall impression (sentiment) and was utilized for comparing the different user groups (low vs. high trust) in further analysis.

We conducted a chi-square test to determine whether there was a significant difference between the two groups (low trust vs. high trust) across the seven aforementioned categories (useful, automated, financial savings, artificial, uncanny, manipulative, dangerous). The percentage frequency of term occurrences for the categories useful, $X^2(1,96) = 11.068$, $p < 0.001$, φ = 0.34, and dangerous, $X^2(1,96) = 7.634$, $p = 0.006$, φ = −0.28, show significant differences between the low trust and high trust group. People who trusted the AI banking assistant perceived fewer risks and danger (5.7%) in using it, as opposed to the user group that distrusts the assistant, and thus used more terms to describe the potential dangers of its use (29.5%; e.g., "deceptive", "suspect"). Similarly, the group that trusted the AI banking assistant regarded it as useful (45.7%; e.g., "optimization", "workload reduction") whereas a substantial number of those who distrusted the AI banking assistant viewed its assistance as less useful (14.8%) and, as previously stated, risky and dangerous (29.5%).

In the remaining five categories, no statistically significant differences can be found, whereby three of them reach a borderline area on an alpha level of 0.05: Automated, $X^2(1,96) = 2.942$, $p = 0.086$, φ = 0.175; financial savings, $X^2(1,96) = 3.749$, $p = 0.053$, φ = 0.198; artificial, $X^2(1,96) = 3.286$, $p = 0.070$, φ = −0.185; uncanny, $X^2(1,96) = 2.079$, $p = 0.149$, φ = −0.147; manipulative, $X^2(1,96) = 0.617$, $p = 0.432$, φ = −0.080. Approximately 20% of the participants belonging to the low trust group perceived the AI banking assistant as uncanny, and around 34% associated terms such as "artificial" with the assistant. Conversely, in the high trust group, the number of participants who described the AI banking assistant as scary is just under 9% and only around 17% described the assistant as artificial. Furthermore, in the high trust group, ∼31%—more than twice the number in the low trust group (15%)—associated the assistant with financial savings.

Comparing the mean sentiment score (overall impression) between the low trust ($M = -0.48$) and high trust ($M = 0.39$) groups also showed a significant difference. Specifically, participants who trusted the AI banking assistant exhibited a significantly more positive attitude (e.g., "practical", "expectant") toward the assistant compared to those who distrusted it. Conversely, the latter group demonstrated a negative attitude, as evidenced by their use of negative associations and terms toward the AI banking assistant (e.g., "unnecessary", "dubious").

3.2. Analysis of closed-ended scales

In the subsequent analysis, we compared the two user groups low trust vs. high trust based on the scale indices of the variables. Wilcoxon-Mann-Whitney tests revealed significant differences in mean values for the following variables: Trust ($U = 0.00$, $Z = -8.378$, $p < 0.001$, $r = 0.74$), intention to use ($U = 353.50$, $Z = -5.687$, $p < 0.001$, $r = 0.50$), competence ($U = 323.00$, $Z = -5.880$, $p < 0.001$, $r = 0.52$), understandability ($U = 313.50$, $Z = -6.016$, $p < 0.001$, $r = 0.53$), intention of developers ($U = 318.50$, $Z = -5.959$, $p < 0.001$, $r = 0.53$), human-likeness ($U = 631.00$, $Z = -3.592$, $p < 0.001$, $r = 0.32$), and uncanniness ($U = 615.50$, $Z = -3.713$, $p < 0.001$, $r = 0.33$). Except for the variable uncanniness, the participants from the high trust group show higher mean values in all variables. In the variable uncanniness, the mean value from the low trust group (M = 3.52, SD = 0.85) is significantly higher compared to the high trust users (M = 2.71, SD = 0.96). All means and standard deviations including all participants ($N = 127$) can be found in Appendix Table 1. All means calculated separately for the two user groups (high trust vs. low trust) are shown in Figure 2.

3.3. Analysis of individual differences

There was no significant difference observed between the low trust and high trust user groups regarding the variables neuroticism ($U = 965.50$, $Z = -1.118$, $p = 0.264$, $r = 0.10$) and openness to experience ($U = 1,108.50$, $Z = -0.056$, $p = 0.955$, $r = 0.05$). The personality trait propensity to trust ($U = 546.50$, $Z = -4.264$, $p < 0.001$, $r = 0.38$), on the other hand, differed significantly between the two groups. Specifically, participants who displayed a high level of trust in the financial assistant also had a higher tendency to trust in general, as compared to those who displayed low levels of trust in the AI banking assistant. No significant differences were found between the low trust vs. high trust group in terms of the

distribution of age ($U = 972.00$, $Z = -1.064$, $p = 0.288$, $r = 0.09$) and gender ($U = 875.00$, $Z = -1.559$, $p = 0.119$, $r = 0.13$).

3.4. Analysis of group differences with regard to agency (manipulation)

To verify the effectiveness of our manipulation of the two conditions (high/low), participants were asked to rate the level of autonomy exhibited by the AI banking assistant using the agency scale. We computed the mean values and detected significant differences between the groups ($U = 687.00$, $Z = -6.451$, $p < 0.001$, $r = 0.57$), indicating a successful manipulation of the conditions. The high agency bot was perceived as a highly autonomous assistant, while the low agency bot was perceived as a less autonomous assistant.

However, a subsequent analysis investigating differences in levels of trust ($U = 1,943.00$, $Z = -0.343$, $p = 0.732$, $r = 0.03$) and intention to use ($U = 1,943.50$, $Z = -0.339$, $p = 0.735$, $r = 0.03$) showed no significant differences between the low and high agency condition. Furthermore, no significant differences were observed between the low and high agency conditions in terms of the other variables, competence ($U = 1,822.00$, $Z = -0.928$, $p = 0.353$), understandability ($U = 1,860.00$, $Z = -0.753$, $p = 0.452$, $r = 0.07$), intention of developers ($U = 1,973.00$, $Z = -0.196$, $p = 0.845$, $r = 0.02$), human-likeness ($U = 1,922.50$, $Z = -0.439$, $p = 0.661$, $r = 0.04$), and uncanniness ($U = 1,722.00$, $Z = -1.414$, $p = 0.157$, $r = 0.13$).

4. Discussion

The findings indicate significant differences between the two groups of users, low trust vs. high trust, which we discovered through open-ended and closed-ended questionnaires. The analysis revealed that the perceptions and attitudes toward the AI banking assistant varied significantly based on the user groups, regarding its potential benefits (useful) and the perceived possibility of harm (dangerous). Specifically, the high trust group acknowledged the financial assistant's usefulness and viewed it as a valuable tool that could assist with tasks, while being less worried about the potential risks associated with it. In contrast, individuals who were skeptical of the AI banking assistant perceived only little assistance from the bot and instead saw the significant dangers that it could pose. In examining the other categories, such as manipulation and uncanniness, there were no significant differences in the frequency of terms between the two user groups. However, the participants of the low trust group perceived the AI banking assistant as particularly scary (almost 20%). Additionally, 16% of the users of the low trust group perceived the AI banking assistant as automated and approximately 34% as artificial.

Consistent with the literature on trust in automation (Körber, 2019), competence was found to be essential and has a significant impact on the level of trust among all participants. Additionally, usefulness, as one of the two key components in the Technology Acceptance Model (Venkatesh and Bala, 2008), proves to be a determining factor in influencing the intention to use the AI banking assistant, as evidenced by the difference between the two

user groups. The strong correlations between competence and trust are apparent in the z-order correlations. To increase trust, one could highlight the system's competence to the low trust group. Additionally, a developer's benevolent intentions can positively impact the level of trust and play a crucial role in the trust calibration process.

The analysis of the closed-ended scales revealed significant differences between the two user groups in all dependent variables. However, unlike the open-ended comments, those who had a lack of trust in the AI system, perceived it as more uncanny and less human-like. Low trust users perceived the AI banking assistant as less competent and the developers' intentions as less benevolent. They also had a lower understanding of the system's workings and calculations, resulting in correspondingly lower intentions to use the tool. Conversely, in the high trust group, all variables, except uncanniness, showed significantly higher values. The AI banking assistant was perceived as highly competent, and users believed that the developers had benevolent intentions. Additionally, users largely understood how the assistant made decisions and took appropriate actions, which increased trust and contributed to a high intention to use the tool.

To establish an effective trust calibration, it is important to address the concerns of both user groups and adjust the level of trust accordingly (Lee and Moray, 1992; Boyce et al., 2015). For the low trust group, it would be beneficial to provide detailed information about the assistant's functions, workings, and calculations to enhance understandability, which directly correlates with trust. By providing information on how the system makes recommendations and decisions, the "black-box" phenomenon could be avoided (e.g., Guidotti et al., 2019). Conversely, the high trust group may be at risk of over-trusting and underestimating the potential dangers of AI powered tools, including autonomous driving or other high-risk applications. Here, clarifying possible risks could facilitate necessary trust calibration to prevent the use of overly risky tools.

In this context, explainable AI (XAI) plays a major role in developing AI systems that provide understandable explanations for their output. Considering these objectives, AI systems should become more transparent and interpretable to users, which is particularly crucial in critical applications where trust, accountability, and adherence to regulations are indispensable (Ehsan et al., 2021). XAI methodologies are typically applied to enhance the interpretability of black-box models for lay human users and to allow them to better assess the trustworthiness of a system or its output (Guidotti et al., 2019; Alicioglu and Sun, 2022; Leichtmann et al., 2023a,b). Although there are models that possess inherent explainability (i.e., models characterized by a significant level of transparency; e.g., Barredo Arrieta et al., 2020), they frequently exhibit the drawback of yielding less precise outcomes. Consequently, *post-hoc* explanation techniques are harnessed to expound upon extant models (e.g., Dosilovic et al., 2018). According to this, requirements collectively aim to strike, among other things, a balance between transparency (AI models' decision-making processes and internal mechanisms), accuracy and performance (not compromise the accuracy and performance of the AI models), and user-centricity (explanations generated should be tailored to the cognitive abilities and needs of the target audience), thereby fostering trust and

confidence in AI technologies (Dosilovic et al., 2018; Barredo Arrieta et al., 2020). Modern emotional AI, by recognizing and adapting to user emotions (Lausen and Hammerschmidt, 2020; Li and Deng, 2022), enhances AI system responses for personalized interactions, may foster higher levels of trust through empathetic and contextually aware communication. This may also contribute to improved user satisfaction, engagement, and long-term reliability in human-machine interactions (e.g., Vishwakarma et al., 2021), but it could also generate skepticism or even fuel overtrust in AI-powered tools (e.g., Grill and Andalibi, 2022).

It is interesting that there were no differences in the variables based on the level of agency of the AI banking assistant (manipulation). This suggests that whether the assistant acted more or less autonomously did not matter. Contrary to our expectations based on the uncanny valley hypothesis (Gray and Wegner, 2012; Mori et al., 2012; Appel et al., 2020), the degree of autonomy, especially in the field of banking, seems to have no impact on the sense of uncanniness experienced. This could be attributed to the fact that our study did not depict a thinking robot capable of generating own ideas, possessing desires, or expressing needs autonomously. Instead, it presented an AI banking assistant as a tool utilizing mathematical calculations to provide savings recommendations and perform other "factual" tasks, devoid of emotional elements. Whereas, what mattered was whether the bot was generally perceived as useful or potentially hazardous. One might expect that individuals with higher levels of neuroticism would perceive greater risks associated with the bot, leading to higher neuroticism scores in the low trust group. However, this was not observed, and the distribution of openness to experience scores between the two user groups was the same. In contrast, the personality trait propensity to trust shows a significant difference between the two user groups. The propensity to trust score is significantly higher in the high trust group, suggesting that trust calibration efforts should be tailored to individual user groups. Consistent with previous research, it is important to consider implications for transparency when developing different AI powered tools for various personalities and user groups (Chien et al., 2016). The advantageous use of an AI powered system could be emphasized and providing information could be used to mitigate unfounded risks and dangers.

Distinctions between users who trust and those who distrust a particular system can manifest in various ways, including behaviors, attitudes, and perceptions. It seems that individuals who trust tend to have a more positive view on AI powered tools and are more likely to engage with the system. On the other hand, those who distrust may exhibit negative attitudes and behaviors toward the system, including avoidance or even actively working against it (e.g., "just useless", "attempted fraud"). Another distinction is that individuals who distrust may be more critical of the system, scrutinizing its actions and decisions more closely. They may also be more hesitant to share personal information or participate in activities that involve the system. In contrast, those who trust may be more willing to provide personal information and engage in activities that involve the system, believing that their information will be handled responsibly. Overall, these differences between the two user groups who trust, and distrust

can have significant implications for the success and effectiveness of the system in question. Understanding these distinctions and working to address them can help build trust and improve user experiences.

5. Limitations and outlook

In terms of trust calibration, it has been found that the two distinct groups of low trust and high trust are appropriate. However, in a larger sample, it would be beneficial to further investigate the individual differences and personality traits among the participants to adopt a more user-centered approach. The present study did not document any pre-experience with an AI banking assistant or AI powered technology in general, which could potentially impact the development of trust. In subsequent surveys, greater emphasis and analysis should be placed on financial literacy. Since the spontaneous reactions (open-ended descriptions) were only asked at the end of the questionnaire, this represents a further limitation and should be considered in subsequent studies. The study was conducted within the context of financial savings, which, in line with classic trust theories, entails potential risks and dangers. Of course, there are numerous other contexts in which similar conditions apply, and it would therefore be worthwhile to examine the different approaches of user groups in other contexts, such as autonomous driving, more closely.

Various AI-powered tools utilize statistical models to analyze extensive data, learning patterns and connections among words and phrases, with the goal of simulating human intelligence and behavior (Bostrom and Yudkowsky, 2014; Hanson Robotics, 2023; Kosinski, 2023). The choice of these tools, like deep learning or genetic algorithms, may also have the potential to shape perceptions and impressions people have, thereby influencing trust (e.g., Wang et al., 2023). Notably, prominent language models like OpenAI's ChatGPT, Meta's LLaMA, or Google's PaLM2 may vary in attributes like fairness, accuracy, or reliability (Chang et al., 2023; Kaddour et al., 2023; Open AI, 2023). The quality of AI inference, referring to a model's capability to generate predictions or responses based on context and input, may also impact users' reliance on AI (e.g., Toreini et al., 2020). Although this aspect wasn't explored in our current study, it presents a potential avenue for future investigations into the alignment of objective and subjective trustworthiness. Establishing trust hinges on AI producing reliable and unbiased outcomes, particularly in collaborative human-AI decision contexts. Designing AI systems for accuracy, fairness, and transparency thus contributes to strengthening user confidence.

6. Conclusion

In this study, we explored initial attitudes and perceptions toward an AI-powered financial assistant, differentiating between two distinct user groups: Those who, after only brief exposure to the system, indicate that they trust it and those who state that they do not. Our findings highlight significant variations in group characteristics and user perceptions. While high trust users acknowledged the AI assistant's usefulness for banking tasks and were less concerned about associated risks, low trust users perceived

limited assistance value, found the system uncanny and difficult to understand, and emphasized potential dangers. These insights underscore the importance of recognizing diverse user groups and trust levels in the development of AI-powered tools for banking and beyond. By tailoring trust calibration strategies to different first impressions of the same system, a potential underestimation of risks among users with high initial trust can be more effectively addressed, while unwarranted distrust in reliable systems can also be countered.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

This research complied with the tenets of the Declaration of Helsinki, the ethical guidelines of the APA Code of Conduct, local legislation, and institutional requirements. Digital informed consent was obtained from each participant.

Author contributions

SS organized the database, performed the statistical analysis, and wrote the first draft of the manuscript. All authors contributed to conception and design of the study, manuscript revision, read, and approved the submitted version.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2023.1241290/full#supplementary-material

References

Alicioglu, G., and Sun, B. (2022). A survey of visual analytics for Explainable Artificial Intelligence methods. *Comp. Graph.* 102, 502–520. doi: 10.1016/j.cag.2021.09.002

Appel, M., Izydorczyk, D., Weber, S., Mara, M., and Lischetzke, T. (2020). The uncanny of mind in a machine: humanoid robots as tools, agents, and experiencers. *Comput. Hum. Behav.* 102, 274–286. doi: 10.1016/j.chb.2019.07.031

Azhikodan, A. R., Bhat, A. G. K., and Jadhav, M. V. (2019). "Stock trading bot using deep reinforcement learning," in *Innovations in Computer Science and Engineering (Bd. 32, S. 41–49),* eds H. S. Saini, R. Sayal, A. Govardhan, R. Buyya (Singapore: Springer Singapore).

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., et al. (2020). Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inform. Fusion* 58, 82–115. doi: 10.1016/j.inffus.2019.12.012

Beer, J. M., Fisk, A. D., and Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *J. Hum. Robot Int.* 3, 74. doi: 10.5898/JHRI.3.2.Beer

Bostrom, N., and Yudkowsky, E. (2014). "The ethics of artificial intelligence,". in *The Cambridge Handbook of Artificial Intelligence.* 1st ed, eds K. Frankish, W. M. Ramsey (Cambridge: Cambridge University Press), 316–334.

Boyce, M. W., Chen, J. Y. C., Selkowitz, A. R., and Lakhmani, S. G. (2015). "Effects of agent transparency on operator trust," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts* (New York, NY: ACM), 179–180.

Burbach, L., Halbach, P., Plettenberg, N., Nakayama, J., Ziefle, M., and Calero Valdez, A. (2019)."Hey, Siri", "Ok, Google", "Alexa". acceptance-relevant factors of virtual voice-assistants," in *2019 IEEE International Professional Communication Conference (ProComm)* (New York City, NY: IEEE), 101–111.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., et al. (2023). *A Survey on Evaluation of Large Language Models.* Ithaca; New York, NY: Cornell University.

Chaves, A. P., and Gerosa, M. A. (2021). How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *Int. J. Hum. Comp. Int.* 37, 729–758. doi: 10.1080/10447318.2020.1841438

Chen, C. C., Chen, X.-P., and Meindl, J. R. (1998). How can cooperation be fostered? The cultural effects of individualism-collectivism. *Acad. Manag. Rev.* 23, 285. doi: 10.2307/259375

Chen, J. Y. C., and Terrence, P. I. (2009). Effects of imperfect automation and individual differences on concurrent performance of military and robotics tasks in a simulated multitasking environment. *Ergonomics* 52, 907–920. doi: 10.1080/00140130802680773

Chien, S.-Y., Lewis, M., Sycara, K., Jyi-Shane, L.iu, and Kumru, A. (2016). "Influence of cultural factors in dynamic trust in automation," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (New York City, NY: IEEE), 002884–002889.

Chong, L., Zhang, G., Goucher-Lambert, K., Kotovsky, K., and Cagan, J. (2022). Human confidence in artificial intelligence and in themselves: the evolution and impact of confidence on adoption of AI advice. *Comput. Human Behav.* 127, 107018. doi: 10.1016/j.chb.2021.107018

Chowdhury, S., Dey, P., Joel-Edgar, S., Bhattacharya, S., Rodriguez-Espindola, O., Abadie, A., et al. (2023). Unlocking the value of artificial intelligence in human resource management through AI capability framework. *Hum. Resour. Manag. Rev.* 33, 100899. doi: 10.1016/j.hrmr.2022.100899

Christoforakos, L., Gallucci, A., Surmava-Große, T., Ullrich, D., and Diefenbach, S. (2021). Can robots earn our trust the same way humans do? A systematic exploration of competence, warmth, and anthropomorphism as determinants of trust development in HRI. *Front. Robot. AI* 8, 640444. doi: 10.3389/frobt.2021.640444

Cohen, J. (1992). Statistical power analysis. *Curr. Dir. Psychol. Sci.* 1, 98–101. doi: 10.1111/1467-8721.ep10768783

Colwell, S. R., and Hogarth-Scott, S. (2004). The effect of cognitive trust on hostage relationships. *J. Serv. Market.* 18, 384–394. doi: 10.1108/08876040410548302

Das, S., Biswas, S., Paul, A., and Dey, A. (2018). "AI doctor: an intelligent approach for medical diagnosis," in *Industry Interactive Innovations in Science, Engineering and Technology,* eds S. Bhattacharyya, S. Sen, M. Dutta, P. Biswas, H. Chattopadhyay (Singapore: Springer Singapore), 173–183.

Day, M.-Y., Cheng, T.-K., and Li, J.-G. (2018). "AI robo-advisor with big data analytics for financial services," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (New York City, NY: IEEE), 1027–1031.

de Visser, E. J., Peeters, M. M. M., Jung, M. F., Kohn, S., Shaw, T. H., Pak, R., et al. (2020). Towards a theory of longitudinal trust calibration in human–robot teams. *Int. J. Soc. Robot.* 12, 459–478. doi: 10.1007/s12369-019-00596-x

Devlin, J. F., Ennew, C. T., Sekhon, H. S., and Roy, S. K. (2015). Trust in financial services: retrospect and prospect. *J. Fin. Serv. Mark.* 20, 234–245. doi: 10.1057/fsm.2015.21

Dilsizian, S. E., and Siegel, E. L. (2014). Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. *Curr. Cardiol. Rep.* 16, 441. doi: 10.1007/s11886-013-0441-8

Dosilovic, F. K., Brcic, M., and Hlupic, N. (2018). "Explainable artificial intelligence: a survey," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (Rijeka: Croatian Society for Information and Communication Technology, Electronics and Microelectronics), 210–215.

Ehsan, U., Wintersberger, P., Liao, Q. V., Mara, M., Streit, M., Wachter, S., et al. (2021). "Operationalizing human-centered perspectives in explainable AI," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (New York, NY: ACM), 1–6.

Eigenstetter, M. (2020). "Ensuring trust in and acceptance of digitalization and automation: contributions of human factors and ethics," in *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Communication, Organization and Work,* ed V. G. Duffy (New York, NY: Springer International Publishing), 254–266.

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Fein, M. L. (2015). Robo-advisors: a closer look. *SSRN Elect. J.* doi: 10.2139/ssrn.2658701

Flores, F., and Solomon, R. C. (1998). Creating trust. *Bus. Ethics Q.* 8, 205–232. doi: 10.2307/3857326

Fox, J., Ahn, S. J., Janssen, J. H., Yeykelis, L., Segovia, K. Y., and Bailenson, J. N. (2015). Avatars versus agents: a meta-analysis quantifying the effect of agency on social influence. *Hum. Comp. Int.* 30, 401–432. doi: 10.1080/07370024.2014.921494

Freedy, A., DeVisser, E., Weltman, G., and Coeyman, N. (2007). "Measurement of trust in human-robot collaboration," in *2007 International Symposium on Collaborative Technologies and Systems* (New York City, NY: IEEE), 106–114.

Gray, H. M., Gray, K., and Wegner, D. M. (2007). Dimensions of mind perception. *Science* 315, 619–619. doi: 10.1126/science.1134475

Gray, K., and Wegner, D. M. (2012). Feeling robots and human zombies: mind perception and the uncanny valley. *Cognition* 125, 125–130. doi: 10.1016/j.cognition.2012.06.007

Grill, G., and Andalibi, N. (2022). Attitudes and folk theories of data subjects on transparency and accuracy in emotion recognition. *Proc. ACM Hum. Comp. Int.* 6, 1–35. doi: 10.1145/3512925

Grounds, C. B., and Ensing, A. R. (2000). Automation distrust impacts on command and control decision time. *Proc. Hum. Fact. Ergon. Soc. Annual Meet.* 44, 855–858. doi: 10.1177/154193120004402293

Guadagno, R. E., Blascovich, J., Bailenson, J. N., and McCall, C. (2007). Virtual humans and persuasion: the effects of agency and behavioral realism. *Media Psychol.* 10, 1–22.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2019). A survey of methods for explaining black box models. *ACM Comp. Surv.* 51, 1–42. doi: 10.1145/3236009

Hannibal, G. (2023). *The Trust-Vulnerability Relation-A Theory-driven and Multidisciplinary Approach to the Study of Interpersonal Trust in Human-Robot Interaction.* Vienna: Technische Universität Wien.

Hanson Robotics (2023). Sophia. Available online at: http://www.hansonrobotics.com (accessed March 27, 2023).

Hayashi, Y., and Wakabayashi, K. (2017). "Can AI become reliable source to support human decision making in a court scene?" in *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing,* 195–198.

He, G., Kuiper, L., and Gadiraju, U. (2023). "Knowing About Knowing: An Illusion of Human Competence Can Hinder Appropriate Reliance on AI Systems," in *CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (New York, NY: ACM).

Ho, C.-C., and MacDorman, K. F. (2010). Revisiting the uncanny valley theory: developing and validating an alternative to the Godspeed indices. *Comput. Human Behav.* 26, 1508–1518. doi: 10.1016/j.chb.2010.05.015

Johnson, D., and Grayson, K. (2005). Cognitive and affective trust in service relationships. *J. Bus. Res.* 58, 500–507. doi: 10.1016/S0148-2963(03)00140-1

Jung, D., Dorner, V., Glaser, F., and Morana, S. (2018). Robo-advisory: digitalization and automation of financial advisory. *Bus. Inform. Syst. Eng.* 60, 81–86. doi: 10.1007/s12599-018-0521-9

Kaddour, J., Harris, J., Mozes, M., Bradley, H., Raileanu, R., and McHardy, R. (2023). *Challenges and Applications of Large Language Models.* Ithaca, New York, NY: Cornell University.

Kamaraj, A. V., and Lee, J. D. (2022). Accumulating distrust: a dynamic model of automation disuse. *Proc. Hum. Fact. Ergonom. Soc. Annual Meet.* 66, 1315–1319. doi: 10.1177/1071181322661380

Kang, H., and Lou, C. (2022). AI agency vs. human agency: understanding human–AI interactions on TikTok and their implications for user engagement. *J. Comp. Med. Commun.* 27, zmac014. doi: 10.1093/jcmc/zmac014

Kim, D. (2005). *Cognition-Based Versus Affect-Based Trust Determinants in E-Commerce: Cross-Cultural Comparison Study. Twenty-Sixth International Conference on Information Systems.* Las Vegas, NV: AIS.

Kochhar, K., Purohit, H., and Chutani, R. (2019). "The rise of artificial intelligence in banking sector," in *The 5th International Conference on Educational Research and Practice (ICERP) 2019* (Universiti Putra Malaysia (UPM), Selangor: The Faculty of Educational Studies), 127.

Kohn, S. C., de Visser, E. J., Wiese, E., Lee, Y.-C., and Shaw, T. H. (2021). Measurement of trust in automation: a narrative review and reference guide. *Front. Psychol.* 12, 604977. doi: 10.3389/fpsyg.2021.604977

Körber, M. (2019). "Theoretical considerations and development of a questionnaire to measure trust in automation," in *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018),* eds, Bagnara, S., Tartaglia, R., Albolino, S., Alexander, T., Fujita, Y., (New York, NY: Springer International Publishing), 13–30.

Körber, M., Baseler, E., and Bengler, K. (2018). Introduction matters: manipulating trust in automation and reliance in automated driving. *Appl. Ergon.* 66, 18–31. doi: 10.1016/j.apergo.2017.07.006

Kosinski, M. (2023). *Theory of Mind May Have Spontaneously Emerged in Large Language Models.* Ithaca; New York, NY: Cornell University.

Kotseruba, I., Gonzalez, O. J. A., and Tsotsos, J. K. (2016). A review of 40 years of cognitive architecture research: focus on perception, attention, learning and applications. *ArXiv Preprint* 1610.08602, 1–74. doi: 10.48550/arXiv.1610.08602

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159. doi: 10.2307/2529310

Lausen, A., and Hammerschmidt, K. (2020). Emotion recognition and confidence ratings predicted by vocal stimulus type and prosodic parameters. *Hum. Soc. Sci. Commun.* 7, 2. doi: 10.1057/s41599-020-0499-z

Lee, J., and Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics* 35, 1243–1270. doi: 10.1080/00140139208967392

Lee, J. D., and See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Hum. Fact. J. Hum. Fact. Ergon. Soc.* 46, 50–80. doi: 10.1518/hfes.46.1.50.30392

Leichtmann, B., Humer, C., Hinterreiter, A., Streit, M., and Mara, M. (2023b). Effects of Explainable Artificial Intelligence on trust and human behavior in a high-risk decision task. *Comput. Human Behav.* 139, 107539. doi: 10.1016/j.chb.2022.107539

Leichtmann, B., Hinterreiter, A., Humer, C., Streit, M., and Mara, M. (2023a). Explainable Artificial Intelligence improves human decision-making: results from a mushroom picking experiment at a public art festival. *Int. J. Hum. Comp. Int.* 1–18. doi: 10.1080/10447318.2023.2221605

Lele, S., Gangar, K., Daftary, H., and Dharkar, D. (2020). Stock market trading agent using on-policy reinforcement learning algorithms. *SSRN Elect. J.* doi: 10.2139/ssrn.3582014

Li, S., and Deng, W. (2022). Deep facial expression recognition: a survey. *IEEE Trans. Affect. Comp.* 13, 1195–1215. doi: 10.1109/TAFFC.2020.2981446

Lieto, A., Bhatt, M., Oltramari, A., and Vernon, D. (2018). The role of cognitive architectures in general artificial intelligence. *Cogn. Syst. Res.* 48, 1–3. doi: 10.1016/j.cogsys.2017.08.003

Ludden, C., Thompson, K., and Moshin, I. (2015). The rise of robo-advice: changing the concept of wealth management. Available online at: https://www.accenture.com/at-de (accessed May 15, 2023).

Lyon, F., Möllering, G., and Saunders, M. (2015). *Handbook of Research Methods on Trust.* Cheltenham: Edward Elgar Publishing.

MacDorman, K. F., and Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. Interaction Studies. *Soc, Behav. Commun. Biol. Artific. Syst.* 7, 297–337. doi: 10.1075/is.7.3.03mac

MacKenzie, D. (2005). Opening the black boxes of global finance. *Rev. Int. Pol. Econ.* 12, 555–576. doi: 10.1080/09692290500240222

Manikonda, L., Deotale, A., and Kambhampati, S. (2018). "What's up with privacy?: User preferences and privacy concerns in intelligent personal assistants," in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 229–235.

Matthews, G., Lin, J., Panganiban, A. R., and Long, M. D. (2020). Individual differences in trust in autonomous robots: implications for transparency. *IEEE Trans. Hum. Mach. Syst.* 50, 234–244. doi: 10.1109/THMS.2019.2947592

Mayer, R. C., Davis, J. H., and Schoorman, F. D. (1995). An integrative model of organizational trust. *Acad. Manag. Rev.* 20, 709. doi: 10.2307/258792

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Acad. Manag. J.* 38, 24–59. doi: 10.2307/256727

McGinn, C., and Torre, I. (2019). "Can you tell the robot by the voice? an exploratory study on the role of voice in the perception of robots," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 211–221.

Medhi Thies, I., Menon, N., Magapu, S., Subramony, M., and O'Neill, J. (2017). "How do you want your chatbot? An exploratory wizard-of-oz study with young, Urban Indians," in *Human-Computer Interaction—INTERACT 2017*, eds R. Bernhaupt, G. Dalvi, A. Joshi, D. K. Balkrishan, J. O'Neill, M. Wincklerm (New York, NY: Springer International Publishing), 441–459.

Miller, D., Johns, M., Mok, B., Gowda, N., Sirkin, D., Lee, K., et al. (2016). Behavioral measurement of trust in automation: the trust fall. *Proc. Hum. Fact. Ergon. Soc. Annual Meet.* 60, 1849–1853. doi: 10.1177/1541931213601422

Miller, T. (2019). Explanation in artificial intelligence: insights from the social sciences. *Artif. Intell.* 267, 1–38. doi: 10.1016/j.artint.2018.07.007

Moradbakhti, L., Schreibelmayr, S., and Mara, M. (2022). Do men have no need for "Feminist," artificial intelligence? Agentic and gendered voice assistants in the light of basic psychological needs. *Front. Psychol.* 13, 855091. doi: 10.3389/fpsyg.2022.855091

Mori, M. (1970). Bukimi no tani [The uncanny valley]. *Energy* . 7, 33–35.

Mori, M., MacDorman, K., and Kageki, N. (2012). The uncanny valley [From the Field]. *IEEE Robot. Automat. Magaz.* 19, 98–100. doi: 10.1109/MRA.2012.2192811

Nass, C. I., Moon, Y., Kim, E., and Fogg, B. J. (1997). *Computers are Social Actors: A Review of Current Research. In Moral and Ethical Issues in Human-Computer Interaction.* Boston, MA: The MIT Press.

Nass, C., and Brave, S. (2005). *Wired for Speech. How Voice Activates and Advances the Human-Computer Relationship.* Cambridge: MIT Press.

Neser, S. (2016). "Vertrauen," in *Psychologie der Werte.* ed D. Frey (Heidelberg: Springer Berlin Heidelberg), 255–268.

Neyer, F. J., and Asendorpf, J. B. (2018). *Psychologie der Persönlichkeit.* Heidelberg: Springer Berlin Heidelberg.

Olson, K. E., Fisk, A. D., and Rogers, W. A. (2009). Collaborative automated systems: older adults' mental model acquisition and trust in automation. *Proc. Hum. Fact. Ergon. Soc. Annual Meet.* 53, 1704–1708. doi: 10.1177/154193120905302210

Open AI (2023). ChatGPD. Computer software. Available online at: https://openai.com/blog/chatgpt (accessed June 01, 2023).

Parasuraman, R., and Riley, V. (1997). Humans and automation: use, misuse, disuse, abuse. *Hum. Fact. J. Hum. Fact. Ergon. Soc.* 39, 230–253. doi: 10.1518/001872097778543886

Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: viable, empirically

supported cognitive engineering constructs. *J. Cogn. Eng. Dec. Mak.* 2, 140–160. doi: 10.1518/155534308X284417

Parviainen, J., and Coeckelbergh, M. (2021). The political choreography of the Sophia robot: beyond robot rights and citizenship to political performances for the social robotics market. *AI Soc.* 36, 715–724. doi: 10.1007/s00146-020-01104-w

Patent, V., and Searle, R. H. (2019). Qualitative meta-analysis of propensity to trust measurement. *J. Trust Res.* 9, 136–163. doi: 10.1080/21515581.2019.1675074

Pitardi, V., Wirtz, J., Paluch, S., and Kunz, W. H. (2021). Service robots, agency, and embarrassing service encounters. *SSRN Elect. J.* 33, 389–414. doi: 10.2139/ssrn.3890941

Pizzi, G., Scarpi, D., and Pantano, E. (2021). Artificial intelligence and the new forms of interaction: who has the control when interacting with a chatbot? *J. Bus. Res.* 129, 878–890. doi: 10.1016/j.jbusres.2020.11.006

Profactor (2021). *Robotische Assistenz.* Available online at: https://www.profactor.at/forschung/industrielle-assistenzsysteme/robotische-assistenz/ (accessed May 15, 2023).

Qiu, R., Ji, Z., Chivarov, N., Arbeiter, G., Weisshardt, F., Rooker, M., et al. (2013). The development of a semi-autonomous framework for personal assistant robots—SRS Project: International Journal of Intelligent Mechatronics and *Robotics*, 3, 30–47. doi: 10.4018/ijimr.2013100102

Questback (2020). Available online at: https://www.questback.com/de/ (accessed June 01, 2023).

Rai, A. (2020). Explainable AI: from black box to glass box. *J. Acad. Market. Sci.* 48, 137–141. doi: 10.1007/s11747-019-00710-5

Rempel, J. K., Holmes, J. G., and Zanna, M. P. (1985). Trust in close relationships. *J. Pers. Soc. Psychol.* 49, 95–112. doi: 10.1037/0022-3514.49.1.95

Righetti, F., and Finkenauer, C. (2011). If you are able to control yourself, I will trust you: the role of perceived self-control in interpersonal trust. *J. Pers. Soc. Psychol.* 100, 874–886. doi: 10.1037/a0021827

Robinette, P., Li, W., Allen, R., Howard, A. M., and Wagner, A. R. (2016). "Overtrust of robots in emergency evacuation scenarios," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 101–108.

Rosenberg, N. (1983). *Inside the Black Box: Technology and Economics* (1. *Aufl.*). Cambridge: Cambridge University Press.

Rosili, N. A. K., Hassan, R., Zakaria, N. H., Kasim, S., Rose, F. Z. C., and Sutikno, T. (2021). A systematic literature review of machine learning methods in predicting court decisions. *IAES Int. J. Artificial Intel.* 10, 1091. doi: 10.11591/ijai.v10.i4.pp1091-1102

Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. *Am. Psychol.* 35, 1–7. doi: 10.1037/0003-066X.35.1.1

Scheuerer-Englisch, H., and Zimmermann, P. (1997). "Vertrauensentwicklung in Kindheit und Jugend," in *Interpersonales Vertrauen,* ed M. Schweer (Wiesbaden: VS Verlag für Sozialwissenschaften), 27–48.

Schreibelmayr, S., and Mara, M. (2022). Robot voices in daily life: vocal human-likeness and application context as determinants of user acceptance. *Front. Psychol.* 13, 787499. doi: 10.3389/fpsyg.2022.787499

Schupp, J., and Gerlitz, J. Y. (2014). "Big five inventory-soep (bfi-s)," in *Zusammenstellung sozialwissenschaftlicher Items und Skalen* (Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften), 10.

Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., et al. (2020a). Machines as teammates: a research agenda on AI in team collaboration. *Inform. Manag.* 57, 103174. doi: 10.1016/j.im.2019.103174

Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I., and Lowry, P. B. (2020b). Collaborating with technology-based autonomous agents: issues and research opportunities. *Int. Res.* 30, 1–18. doi: 10.1108/INTR-12-2019-0503

Shanmuganathan, M. (2020). Behavioural finance in an era of artificial intelligence: longitudinal case study of robo-advisors in investment decisions. *J. Behav. Exp. Fin.* 27, 100297. doi: 10.1016/j.jbef.2020.100297

Statista (2021). Available online at: https://www.statista.com (accessed March 27, 2023).

Stein, J.-P., Liebold, B., and Ohler, P. (2019). Stay back, clever thing! Linking situational control and human uniqueness concerns to the aversion against autonomous technology. *Comput. Hum. Behav.* 95, 73–82. doi: 10.1016/j.chb.2019.01.021

Stein, J.-P., and Ohler, P. (2017). Venturing into the uncanny valley of mind—the influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition* 160, 43–50. doi: 10.1016/j.cognition.2016.12.010

Tertilt, M., and Scholz, P. (2018). To advise, or not to advise—how robo-advisors evaluate the risk preferences of private investors. *J Wealth Manag.* 21, 70–84. doi: 10.3905/jwm.2018.21.2.070

Tomsett, R., Preece, A., Braines, D., Cerutti, F., Chakraborty, S., Srivastava, M., et al. (2020). Rapid trust calibration through interpretable and uncertainty-aware AI. *Patterns* 1, 100049. doi: 10.1016/j.patter.2020.100049

Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., Zelaya, C. G., and van Moorsel, A. (2020). "The relationship between trust in AI and trustworthy machine learning technologies," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (New York, NY: ACM), 272–283.

Varshney, S. (2016). *Introduction to Trading Bot. In S. Varshney, Building Trading Bots Using Java* (New York, NY: Apress), 1–25.

Venkatesh, V., and Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sci.* 39, 273–315. doi: 10.1111/j.1540-5915.2008.00192.x

Vincent, G., Laknidhi, V., Klein, P., and Gera, R. (2015). Robo-advisors: capitalizing on a growing opportunity. *J. Res. Robo Adv. Technol.*

Vishwakarma, A., Sawant, S., Sawant, P., and Shankarmani, R. (2021). "An emotionally aware friend: moving towards artificial general intelligence," in *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)* (New York City, NY: IEEE), 1094–1100.

Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., and Trichina, E. (2022). Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review. *Int. J. Hum. Res. Manag.* 33, 1237–1266. doi: 10.1080/09585192.2020.1871398

Wang, B., Chen, W., Pei, H., Xie, C., Kang, M., Zhang, C., et al. (2023). *DecodingTrust: A Comprehensive Assessment of Trustworthiness in GPT Models*. Ithaca, New York, NY: Cornell University.

Ward, C., Raue, M., Lee, C., D'Ambrosio, L., and Coughlin, J. F. (2017). "Acceptance of automated driving across generations: the role of risk and benefit perception, knowledge, and trust," in *Human-Computer Interaction. User Interface Design, Development and Multimodality* ed M. Kurosu (New York, NY: Springer International Publishing), 254–266.

Young, M. S., and Stanton, N. A. (2007). Back to the future: brake reaction times for manual and automated vehicles. *Ergonomics* 50, 46–58. doi: 10.1080/00140130600980789

Złotowski, J., Yogeeswaran, K., and Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *Int. J. Hum. Comput. Stud.* 100, 48–54. doi: 10.1016/j.ijhcs.2016.12.008