Journal of Innovation & Knowledge 10 (2025) 100770

Contents lists available at ScienceDirect

Journal of Innovation & Knowledge

journal homepage: www.elsevier.com/locate/jik

The key role of design and transparency in enhancing trust in AI-powered digital agents

Iris Glassberg *, Yael Brender-Ilan, Moti Zwilling *

Deparment of Economics & Business Administration, Ariel University, Ariel 40700, Israel

ARTICLE INFO

JEL classification:
M10
O30
O33
O39

Keywords:
Digital agents
Artificial intelligence
Trust
Expectations
Organizational adoption
ChatGPT

ABSTRACT

This study examines the factors that influence user trust in AI-powered digital agents within organizational contexts. Employing an integrated framework that combines the Expectation Confirmation Model and the Technology Acceptance Model, we investigated how user expectations, attitudes toward AI, information transparency, visual design, and exposure duration contribute to trust formation. A sample of 118 organizational participants in Israel interacted with ChatGPT across simulated workplace scenarios. The findings indicate that while expectations alone do not significantly predict trust, transparency and effective visual design serve as crucial mediators. Attitudes toward AI were also positively associated with trust. The experimental manipulation of expectations did not yield a significant effect, suggesting that pre-existing user perceptions may override brief interventions. These results highlight actionable design and implementation strategies for organizations aiming to foster user trust and promote AI adoption. Limitations include sample bias, constraints in subgroup analysis, and the need for longitudinal research.

Introduction

The digital landscape is undergoing a profound transformation driven by artificial intelligence (AI), reshaping organizational workflows and revolutionizing both personal and professional environments (Diederich et al., 2022; Letheren et al., 2020). The COVID-19 pandemic has accelerated this shift, significantly increasing the adoption of AI-based technologies, particularly digital agents such as chatbots, to support remote and hybrid work environments (Lartey, 2021; Nguyen et al., 2021; Silva et al., 2023).

Digital agents offer numerous advantages: they operate tirelessly, are available 24/7, and can handle multiple requests simultaneously (Lou, 2019; Mohd et al., 2022). These capabilities translate into reduced operational costs and enhanced service efficiency (Blazevic & Sidaoui, 2022; Letheren et al., 2020; Riethorst, 2022). However, one of the primary barriers to adoption remains a lack of user trust. Without trust, users may resist adoption, underutilize the system, or rely on it inadequately. Concerns regarding privacy, data usage, and the opaque "black box" nature of AI algorithms further amplify this skepticism (Grimes et al., 2021; Gulati et al., 2019). Therefore, understanding how trust in AI is established is crucial for the effective and ethical deployment of these technologies. In particular, aligning user expectations, promoting transparency, and designing intuitive user interfaces are likely to play pivotal roles in cultivating trust and fostering long-term engagement.

To address these challenges, this study integrates the Expectation Confirmation Model (ECM) and the Technology Acceptance Model (TAM) to examine the determinants of trust in AI-powered digital agents. While ECM focuses on how users evaluate system performance against their expectations, TAM introduces the concept of perceived ease of use (PEOU), which can influence both initial expectations and the confirmation of those expectations. For example, users may expect more from systems that are easier to use and may attribute imperfections to their errors rather than to the system itself. Integrating these models enables a more nuanced analysis of how cognitive and experiential factors shape trust in AI technologies.

The market for AI-powered assistants is rapidly expanding, with significant investments by technology giants underscoring their transformative potential (Diederich et al., 2022; Terblanche & Kidd, 2022). Despite these benefits, digital agents still face challenges in achieving user acceptance, particularly for complex interactions. Many users prefer human interaction for issues requiring empathy and understanding (Nguyen et al., 2021), highlighting a gap between user expectations and the current capabilities of digital agents. While chatbots are evolving to manage more complex tasks through advancements in

* Corresponding authors.
E-mail addresses: iris.glassber@msmail.ariel.ac.il (I. Glassberg), motiz@ariel.ac.il (M. Zwilling).

https://doi.org/10.1016/j.jik.2025.100770
Received 21 December 2024; Accepted 11 July 2025
Available online 28 July 2025
2444-569X/© 2025 The Author(s). Published by Elsevier España, S.L.U. on behalf of Journal of Innovation & Knowledge. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

natural language processing (NLP), users often perceive them as suitable only for simple queries (Ahonen, 2020; Jiang et al., 2023).

Trust is a critical factor in successfully adopting AI technologies, which is significantly influenced by privacy concerns, lack of transparency, and the "black box" nature of AI systems. User attitudes toward AI can be negatively affected by algorithmic errors and ethical concerns (Gulati et al., 2019). Addressing these trust issues can improve user satisfaction and ensure successful integration.

This study seeks to explore how user expectations influence trust in digital agents within organizational contexts, focusing on ChatGPT as a platform. We examine how factors such as attitudes toward AI, transparency in information usage, agent activation duration, and perceived visual features impact user trust. The research involves 118 participants from Israeli organizations who interacted with ChatGPT in simulated organizational scenarios.

By integrating ECM and TAM and employing experimental manipulation of expectations with real-time interaction, this study advances the understanding of trust in AI. The results offer practical recommendations for organizations and provide actionable insights for enhancing user trust and increasing AI adoption.

By exploring these dynamics, we aim to offer valuable insights for organizations seeking to enhance user trust in their digital agents. Effectively managing user expectations, improving transparency, and designing user-friendly interfaces can foster greater adoption and satisfaction with AI technologies, ultimately leading to improvement in work performance and user experience.

Our study offers a unique exploration of the interaction between organizational users and a virtual agent, utilizing a scenario-based experimental approach. To the best of our knowledge, this is the first investigation to examine such interactions within an organizational context directly, providing new perspectives on the practical integration of AI technologies in the workplace.

Apart from its empirical contribution, this study refines and extends the application of the ECM and TAM by integrating them into a unified framework for understanding trust in AI-driven digital agents and revealing both their explanatory value and theoretical limitations within organizational settings.

Our study presents a novel investigation into how transparency, visual design, and user expectations jointly influence trust in AI-powered agents, using a scenario-based experimental design conducted in a real organizational context. While most prior studies have examined these factors in isolation, few have explored their combined effects—particularly within workplace environments. Furthermore, the integration of ECM and TAM remains underexplored in the literature on AI trust.

By addressing these empirical and theoretical gaps, our study provides new insights into trust formation in organizational AI systems and offers practical implications for the development, design, and deployment of AI-driven digital agents in the workplace.

The remainder of this paper is structured as follows: Section 2 presents the theoretical background and hypothesis development; Section 3 describes the research methodology, including participants, measures, and procedure; Section 4 reports the results of the mediation and moderation analyses; Section 5 discusses the findings referring to prior research; Section 6 outlines theoretical and practical implications; and Section 7 concludes the paper by discussing limitations and directions for future research.

Literature review

Digital conversational agents (Chatbots)

Digital conversational agents, commonly known as chatbots, are computer programs that use NLP to enable users to interact with systems through text or voice, simulating human dialogue (Nordheim, 2018). The origins of natural language interaction in computing date back to the 1960s, with Joseph Weizenbaum's development of "ELIZA," a pioneering chatbot designed to play the role of a psychiatrist (Diederich et al., 2022; Nguyen, 2019). Digital agents can mimic a psychotherapist. Lately, digital agents are widely used in organizations for both external communication, such as customer support, and internal purposes, including employee assistance. Interactions with digital agents can take the form of text-based exchanges, voice-based communication (such as with Amazon's Alexa), or even 3D animated avatars (Ahmad et al., 2022).

These agents enhance organizational productivity by managing various business processes and services at a lower cost, supporting employees in accessing documents, retrieving organizational information, and providing translations (Gkinko & Elbanna, 2022). By filtering information overload and making work resources more accessible, digital agents improve the employee experience and streamline workflows. As AI and machine learning-based tools, digital agents facilitate natural language communication, fostering a sense of human-like interaction (Diederich et al., 2022; Lou, 2019). They are not limited by time or space (Nguyen et al., 2021), allowing organizations to engage with users at any hour and from anywhere across a range of messaging platforms, including Facebook Messenger, Slack, WhatsApp, and WeChat.

The development of chatbots is supported by various platforms, such as Google's DialogFlow, Microsoft Bot Framework, and Pandorabots, as well as open-source frameworks like Rasa and Mycroft. The market for AI-powered assistants is experiencing significant growth, with billions of dollars invested annually (Terblanche & Kidd, 2022). It was estimated that by 2025, 95 % of customer interactions will be managed by AI agents (Nordheim, 2018; Sundareswaran, 2021), and the global AI chatbot market is projected to reach $18.02 billion by 2027 (Gkinko & Elbanna, 2023). Nonetheless, despite these advances, challenges remain. User acceptance remains an issue, as many customers prefer human interaction, especially for complex situations that require empathy and understanding (Nguyen et al., 2021). A 2021 survey found that while 74 % of consumers expect to encounter chatbots on company websites, only 13 % prefer them over human agents.

ChatGPT

ChatGPT, developed by OpenAI, marks a significant advancement in digital conversational agents. Being an advanced NLP model, ChatGPT is capable of adapting to the flow and content (Rathore, 2023) of conversations. Pre-trained on extensive datasets, it can generate coherent and contextually appropriate language that often appears indistinguishable from human-written text. ChatGPT is part of a broader trend in generative AI, which uses trained data to create original content in various formats. This technological evolution has enormous potential for chatbot development, and companies such as Google, Meta, and DeepMind are actively exploring how generative AI can further enhance chatbot capabilities. Google's LaMDA, which powers their Bard chatbot, is a prominent competitor in this space. These advanced models represent a new generation of conversational AI platforms, enabling organizations to build and deploy sophisticated chatbots across multiple communication channels.

Nevertheless, issues of user trust and acceptance persist. Concerns about privacy, a lack of transparency, and the opaque nature of AI systems (Grimes et al., 2021; Kaya et al., 2022) continue to affect user confidence. As these technologies evolve, addressing such trust-related issues becomes crucial for their successful integration into organizational workflows.

Theoretical foundations

Understanding trust in AI-driven digital agents requires a theoretical framework that encompasses both users' pre-adoption expectations and their post-adoption evaluations. To address this, the present study integrates two well-established models: the ECM and TAM. Expectation Confirmation Theory (ECT), introduced by Oliver (1980), explains how satisfaction and subsequent behaviors are shaped by the alignment

between what users expect before using a system and how they perceive its actual performance. The ECM (Bhattacherjee, 2001) extends this idea to information systems, emphasizing how post-adoption experiences influence satisfaction, trust, and continued use. The ECM highlights three key constructs: expectations, confirmation—the extent to which performance meets those expectations—and satisfaction. Regarding AI, such as digital agents like ChatGPT, positive confirmation, where the system's performance meets or exceeds expectations, leads to satisfaction and trust, while negative confirmation results in dissatisfaction and possible rejection (Kosch et al., 2023).

The ECM incorporates concepts from the TAM (Davis, 1989), which focuses on users' initial attitudes and behavioral intentions toward technology. The TAM identifies two central predictors: PEOU, which means how simple the system is to use, and perceived usefulness (PU), in other words, the extent to which the system enhances task performance. In AI-powered systems, intuitive interface design contributes to PEOU, while accurate and efficient task execution enhances PU (Nguyen et al., 2021; Silva et al., 2023).

Integrating ECM and TAM allows for a comprehensive examination of both the pre-adoption and post-adoption dynamics of trust formation. The TAM captures users' initial willingness to engage with AI systems, while ECM explains how ongoing experiences reinforce or undermine trust. The two models are complementary: PU and PEOU shape initial expectations, while confirmation and satisfaction determine whether those expectations ultimately foster sustained trust. Research has shown that aligning expectations with transparent design and effective performance strengthens trust in AI agents (Bhattacherjee, 2001; Shin, 2021).

In organizational settings, this integrated approach helps explain why technically sound AI systems may still fail if institutional transparency and user experience design are neglected (Gkinko & Elbanna, 2023).

Within this framework, trust serves as both an outcome of expectation confirmation and a mediator linking user experience to continued use. For instance, when users find an AI agent's interface intuitive and its performance effective, their expectations are more likely to be confirmed, leading to satisfaction and greater trust. In turn, this trust increases users' willingness to adopt or recommend the system within their organization.

This synthesis addresses a key gap identified in recent systematic reviews (Afroogh et al., 2024) by providing a dynamic model of trust calibration specific to organizational AI adoption. The combined framework offers a more comprehensive understanding of how users form and sustain trust in AI-driven digital agents, bridging pre-use perceptions with post-use evaluations and emphasizing the importance of design, transparency, and user experience in shaping organizational adoption behaviors.

A comparison between ECM and TAM highlights their complementary roles. While TAM focuses on pre-adoption perceptions shaped by design and utility, ECM captures post-adoption evaluations driven by the alignment of expectations and experiences. Together, they provide a lifecycle perspective on trust formation in organizational AI, offering valuable insights into how trust develops and can be supported throughout the adoption process.

Table 1 shows that while TAM focuses on pre-adoption perceptions shaped by design factors such as PEOU and PU, ECM captures post-adoption evaluations driven by the alignment between expectations and actual system performance. Together, these models provide a comprehensive lifecycle perspective on trust formation in organizational AI, highlighting how both initial perceptions and ongoing experiences contribute to the development and maintenance of user trust.

User expectations from digital agents

User expectations play a crucial role in shaping satisfaction and adoption behaviors for new technologies such as digital agents. When interactions with these agents align with users' initial expectations, satisfaction and overall experience tend to be more positive. Conversely, when expectations are unmet, dissatisfaction and even rejection of the technology occur (Kosch et al., 2023).

The ECM (Bhattacherjee, 2001), rooted in ECT (Oliver, 1980) and incorporating constructs from TAM (Davis, 1989), such as PEOU, provides a nuanced framework for understanding how users develop trust in digital agents within organizational contexts.

According to ECT, individuals form expectations about a product or service before use and subsequently evaluate its performance based on their experience. The comparison between expectations and performance either confirms or disconfirms those expectations, directly influencing satisfaction and the intention to continue using the technology. The ECM has proven effective in interpreting continued use intentions for AI-driven service agents (Følstad et al., 2021; Nguyen et al., 2021). Regarding information systems, ECM suggests that user satisfaction is a key determinant of post-adoption behavior, particularly using technology continuously. The literature consistently demonstrates a positive relationship between user satisfaction and the intention to reuse chatbots (Silva et al., 2023).

In particular, expectations are not fixed and can be influenced or manipulated. For example, a study on robot interfaces found that participants preferred interacting with a robot described as feature-rich compared to one described as having fewer features, even when the actual interface was identical. Thus, the way a system is described can shape user expectations and ultimately affect measured outcomes (Kosch et al., 2023). However, managing expectations is particularly challenging in the context of AI, given the extensive media attention and hype around it. Many users develop unrealistic expectations about AI's capabilities, which can lead to distrust and rejection if the technology does not deliver as anticipated (Hinsen et al., 2022). For chatbots, expectations of time-saving, accurate information, and instant support can result in positive experiences if performance meets or exceeds these expectations (Nguyen et al., 2021).

This study examined whether manipulating the description of the digital agent ChatGPT can influence users' expectations and trust, as explored in the experimental component. By understanding and managing user expectations, organizations may improve the adoption and continued use of digital agents across various contexts. Thus, this study aims to determine whether adjusting the description of ChatGPT can impact user expectations and trust, which was assessed in the experiment.

Table 1
Comparison between ECM and TAM.

Aspect	ECM (Bhattacherjee, 2001)	TAM (Davis, 1989)
Focus	Post-adoption evaluation and continuance	Pre-adoption attitudes and intention to use
Key Constructs	Expectations, Confirmation, Satisfaction	Perceived Ease of Use (PEOU), Perceived Usefulness (PU)
Trust Position	Outcome of confirmation and driver of continuance	Shaped by PEOU and PU, influencing intention
Context of Application	Information systems, user experience after adoption	Initial user perceptions and technology adoption
Relevance to AI Agents	Explains sustained trust based on experience	Explains willingness to try based on design/utility

Research hypotheses

The relationship between user expectations of the digital agent and trust in the agent

User expectations of a digital agent can be high or low, and these expectations influence the degree of trust placed in the agent. Previous studies have shown that when users' expectations are met by technology, it positively affects their acceptance and intention to reuse the technology (Hinsen et al., 2022). Accordingly, we hypothesize that greater alignment with user expectations will result in higher trust:

H1: A positive correlation exists between user expectations of the digital agent and the level of trust in the digital agent.

The Moderating Effect of Expectation Setting on the Relationship Between User Expectations and Trust

We anticipate that pre-provided information will shape users' expectations and that the relationship between expectations and trust will be influenced by how expectations are set. Previous research has demonstrated that user expectations can be manipulated and that system descriptions impact both users' expectations and their interactions with AI systems, as well as the metrics assessed (Kosch et al., 2023). We expect that establishing high or realistic expectations will impact the degree of trust in the agent. Therefore, we propose the following hypothesis:

H1a: Expectation setting moderates the relationship between user expectations of the digital agent and the level of trust in the digital agent.

User attitudes toward artificial intelligence (ATAI)

Some individuals are highly receptive to AI products and recognize their advantages, while others are ambivalent or even skeptical, expressing concerns about the potential risks of AI. Notable figures such as Stephen Hawking and Elon Musk have publicly warned about the existential risks posed by AI research (Sindermann et al., 2021). Users often voice concerns that advances in AI could undermine human meaning, productivity, and autonomy, especially by creating human-like thinking machines. Nevertheless, AI offers significant benefits: in cars, it can promote safer driving; in healthcare, it can assist medical professionals; and in daily life, it can simplify tasks through smart devices (Sindermann et al., 2022). Individuals with a more positive attitude toward new technology are more likely to trust chatbots (Ahonen, 2020). Similarly, a positive attitude toward AI is associated with a greater willingness to use AI products, while negative attitudes correlate with lower adoption rates (Sindermann et al., 2021). The "Attitude Toward Artificial Intelligence" (ATAI) scale measures general attitudes toward AI and is correlated with the user's willingness to adopt AI products such as Siri or Alexa (Sindermann et al., 2022).

The Mediating Role of User Attitudes Toward AI in the Relationship Between Expectations and Trust

We hypothesize that users' attitudes toward AI mediate the relationship between their expectations of the interaction and their trust in the agent. A more positive attitude is likely to foster more realistic expectations, which in turn should lead to higher trust. As found in previous studies, a positive attitude toward AI is linked to a greater willingness to use AI products, while negative attitudes are associated with reluctance to adopt (Sindermann et al., 2021).

H2a: A positive correlation is found between users' attitudes toward AI and their expectations of the digital agent.

H2b: A positive correlation is found between users' attitudes toward AI and the level of trust in digital agents.

H2c: Users' attitudes toward AI mediate the relationship between their expectations of the digital agent and the level of trust in the agent.

Disclosure of information usage (Privacy concerns)

The accelerated adoption of conversational agents underscores the importance of safeguarding user information. Users are concerned about the mishandling of sensitive data and fear potential leaks (Pesonen, 2021). AI-based systems can expose users to privacy risks due to the sensitivity of the information involved. Even when technology marketers do not intend to create such issues, privacy concerns can still arise. For example, robotic vacuum cleaners can map users' homes, and unsecured smart home systems can put consumers at risk. Managing these risks is a challenge for companies, which requires them to prevent breaches while anticipating future customer growth in a new and potentially disruptive environment (Grimes et al., 2021). Major technology companies have been involved in privacy breaches, recording and analyzing private conversations through AI products (Sindermann et al., 2021). Companies are responsible for safeguarding privacy, especially when handling sensitive information (Nordheim, 2018). Users may hesitate to share personal information with chatbots if they doubt the security of their data (Silva et al., 2023). Perceptions of security, privacy, and risk also shape users' trust in chatbots. Moreover, if users are unaware that they are interacting with a robot, they may mistakenly believe it is human, leading to surprise or disappointment upon discovering the truth (Nguyen et al., 2021). Regulatory frameworks such as the General Data Protection Regulation GDPR in Europe and Children's Online Privacy Protection Act COPPA in the U.S. require organizations to protect users' data (Mhlanga, 2023). Research demonstrates that transparency, privacy, and data security are variables critical for building trust in digital agents (Gkinko & Elbanna, 2023).

The Mediating Role of Information Disclosure in the Relationship Between Expectations and Trust

We hypothesize that disclosing the non-human nature of the agent, including the extent of data protection, increases the users' confidence that their data is secure, thereby leading to greater trust in the agent.

H3a: A positive correlation exists between the degree of information disclosure and users' expectations of the digital agent.

H3b: A positive correlation exists between the degree of information disclosure and the level of trust in the digital agent.

H3c: The degree of information disclosure mediates the relationship between user expectations of the digital agent and the level of trust in the agent.

Visual design features and PEOU

A third mediating factor is visual design and PEOU. According to TAM, PEOU is a strong predictor of adoption and satisfaction. Visual and interaction design features play a predominant role in how users experience digital agents: intuitive interfaces, clarity of interaction, and responsiveness contribute to a sense of control and competence (Silva et al., 2023). Systems that are visually coherent and simple to operate are more likely to be trusted and perceived as effective. Perceived ease of use (PEOU), as outlined in the TAM, may influence the expectation confirmation process described in the ECM. Specifically, PEOU can affect users' initial expectations of the system, which are perceived as easier to use and may generate higher expectations for performance and reliability. Moreover, PEOU can facilitate the confirmation process itself: if the system is simple and intuitive, users may more easily reconcile minor shortcomings, attributing them to their misunderstanding rather than system failure. Thus, PEOU shapes the formation of expectations and their evaluation after system use. Purposeful user interface design fosters acceptance and trust in AI-based technologies by aligning with user expectations. Acceptance of new technologies is critical to success, and users' satisfaction with the technology determines their intention to continue using the agent (Hinsen et al., 2022). The design refers to the characteristics of the technology used to develop the chatbot and can be categorized into functionality and security. Proper design positively influences user trust, reflecting the system's capabilities (Mohd et al., 2022). Customers are more willing to adopt technology they understand and find easy to use (Silva et al., 2023). Moreover, AI chatbot design features can influence workplace emotions (Gkinko & Elbanna, 2023). If an information system has poor interface design that complicates user interactions, trust in the service provider's ability to deliver high-quality services may be compromised (Nguyen et al., 2021).

The Mediating Role of Visual Features and PEOU in the Relationship Between Expectations and Trust

It is expected that visual features and PEOU simplify interactions and enhance usability, thereby increasing trust in the agent. Thus, we propose the following hypotheses:

H4a: A positive correlation exists between visual features and PEOU and users' expectations of the digital agent.

H4b: A positive correlation exists between visual features and PEOU and the level of trust in the digital agent.

H4c: Visual features and PEOU mediate the relationship between user expectations of the digital agent and the level of trust in the agent.

Duration of exposure to the digital agent

As chatbots and digital agents evolve to handle increasingly complex tasks, users require time to build trust through repeated interactions. While early chatbots were limited to simple functions, advancements in AI have enabled them to perform more sophisticated tasks (Ahonen, 2020). Users are unlikely to fully accept or trust an AI application after just one interaction, particularly when dealing with complex tasks. Instead, trust typically develops over time as users experience successful interactions and the agent consistently meets their expectations (Hinsen et al., 2022). In the initial stages of interaction, users may only explore a subset of the digital agent's capabilities. As they become more familiar with the technology, acceptance tends to grow, especially when the agent reliably fulfills tasks—reinforcing the likelihood of trust in future interactions (Hinsen et al., 2022). This gradual acceptance process is particularly relevant when introducing a novel communication mode such as AI-driven conversation, which often requires multiple interactions for users to feel comfortable and confident (Nordheim, 2018). Over time, as repeated positive interactions occur, the degree of trust stabilizes and becomes a lasting component of the user–agent relationship (Pinto et al., 2022). Moreover, digital agents benefit from ongoing user engagement through continuous interaction. As the agent learns from user behavior, such as language patterns and usage habits, it enhances its ability to respond accurately to user needs, further fostering trust and satisfaction (Gkinko & Elbanna, 2022, 2023). This iterative learning process not only improves the agent's operational capabilities but also creates a more personalized user experience.

The Mediating Role of Duration of Exposure in the Relationship Between Expectations and Trust

Given the mediating effect of exposure duration among user expectations and trust, more exposure allows users to accumulate positive experiences that align with their expectations over time. This cumulative process supports the hypothesis that trust in digital agents increases with repeated interactions (Hinsen et al., 2022). Therefore, we propose the following hypotheses:

H5a: A positive correlation exists between exposure duration to the agent and the users' expectations of the digital agent.

H5b: A positive correlation exists between exposure duration to the agent and the degree of trust in the digital agent.

H5c: The exposure duration to the agent mediates the relationship between user expectations of the digital agent and the degree of trust in the digital agent.

Overall, these hypotheses form an extended version of the ECM adapted to AI-driven digital agents, which integrate mediating psychological, design-related, and behavioral factors to provide a holistic explanation for trust formation in organizational settings.

Degree of trust in digital agents

Trust is a fundamental factor in the successful adoption and sustained use of digital agents within organizations. Research indicates that individuals generally exhibit lower levels of trust in chatbots compared to face-to-face interactions, highlighting the unique challenges digital agents face in establishing credibility (Riethorst, 2022). Trust significantly influences users' willingness to adopt and continue using technology, as it helps reduce perceived risks and uncertainty. When users are confident that digital agents will not misuse their information or exploit vulnerabilities, they are more likely to engage positively with these tools (Silva et al., 2023). Trust is especially crucial for AI-driven chatbots, as it shapes users' perceptions of safety during online interactions and alleviates concerns regarding potential misuse of personal data by malicious actors (Nguyen et al., 2021). Given that chatbots simulate human-like conversations, they inherently carry risks, such as the potential for rogue chatbots designed by hackers to deceive users into divulging sensitive information. Therefore, trust becomes a pivotal factor in user engagement, directly impacting behavioral intentions and influencing long-term adoption. Research shows that users are more inclined to interact with chatbots and digital agents when they trust the technology, which helps mitigate anxieties related to privacy and security (Gkinko & Elbanna, 2023). Consequently, establishing user trust is essential for the widespread adoption and integration of chatbots in organizational contexts, where continuous use heavily relies on trust and perceived reliability (Riethorst, 2022).

The study model

The conceptual model integrates the insights and hypotheses, as illustrated in Fig. 1.

As illustrated in Fig. 1, the conceptual model integrates the key insights and hypotheses of the study, detailing the relationships between user expectations, attitudes toward artificial intelligence (AI), and trust in the digital agent. The figure also emphasizes the moderating and mediating effects of factors such as exposure duration, visual characteristics of the agent, and the extent of information reporting. This framework allows for a systematic analysis of how user-related and agent-specific variables interact to shape trust, offering valuable insights for the design and implementation of trustworthy digital agents in organizational contexts.

Table 2 summarizes the hypotheses presented in the research model (Fig. 1), including their corresponding variables and proposed relationships.

As summarized in Table 2, the hypotheses presented in the research model (Fig. 1) include various independent variables, mediators, and dependent variables. The table outlines the proposed relationships between user expectations, attitudes toward artificial intelligence (AI), visual characteristics, and trust, emphasizing the mediating and moderating effects of factors such as expectation setting, information disclosure, and exposure to visual design features. By mapping these hypothesized pathways, Table 2 enables a systematic analysis of how user-related and agent-specific factors interact - both directly and indirectly - to shape trust in organizational AI systems. This comprehensive mapping of the hypotheses lays a valuable foundation for future empirical analysis, ensuring that all relevant variables and their potential interactions are systematically considered in the study of trust in organizational AI systems.

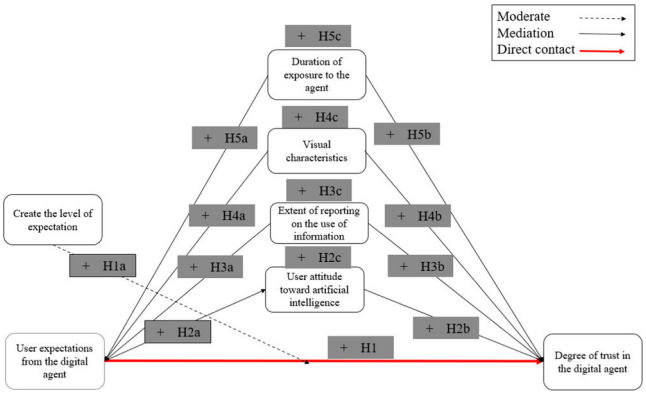


Fig. 1. Research model and hypotheses.

Table 2
The hypotheses.

Hypothesis	Independent Variable	Mediator/Moderator	Dependent Variable
H1	User Expectations	–	Trust
H1a	User Expectations	Expectation Setting	Trust
H2a	Attitude toward AI	–	Expectations
H2b	Attitude toward AI	–	Trust
H2c	Expectations	Attitude toward AI	Trust
H3a	Information Disclosure	–	Expectations
H3b	Information Disclosure	–	Trust
H3c	Expectations	Information Disclosure	Trust
H4a	Visual Characteristics	–	Expectations
H4b	Visual Characteristics	–	Trust
H4c	Expectations	Visual Characteristics	Trust
H5a	Exposure Duration	–	Expectations
H5b	Exposure Duration	–	Trust
H5c	Expectations	Exposure Duration	Trust

Method

Participants

The study sample comprised 118 participants from various organizations in Israel who interacted with ChatGPT and completed an online questionnaire between May and July 2023. Of these participants, 80.5 % ($N = 95$) were male and 19.5 % ($N = 23$) were female. Although the study aimed for a balanced gender distribution, the final sample was predominantly male. Several women who were initially approached declined to participate, citing discomfort or unfamiliarity with AI systems such as ChatGPT. This reluctance may have contributed to the observed gender imbalance. Regarding age, 7.6 % ($N = 9$) were under 30, 55.1 % ($N = 65$) were aged 30–50, and 37.3 % ($N = 44$) were over 50. Educational backgrounds were diverse: 11.9 % ($N = 14$) held no academic degree, 44.9 % ($N = 53$) had a bachelor's degree, 35.6 % ($N = 42$) held a master's degree, and 7.6 % ($N = 9$) had a doctorate or higher. Regarding academic discipline, 40.7 % ($N = 48$) were educated in computer science or technology fields, while 59.3 % ($N = 70$) came from other domains.

Data collection adhered to ethical standards, ensuring participant anonymity and confidentiality. Participation was voluntary and anonymous, and all collected data were kept confidential.

Two types of participants were examined: (1) those with a background in software or technology, knowledgeable about AI; (2) those from other educational backgrounds, less technologically oriented and more ambivalent toward AI. Each participant received a digital kit containing explanatory text and questionnaires, which were subsequently analyzed.

Due to the relatively small sample size, analyses involving subgroups—such as expectation-setting conditions and field of study—were considered exploratory and should be interpreted with caution. These analyses were intended to generate preliminary insights rather than to support definitive conclusions regarding group differences.

Research procedure

To examine the research hypotheses, a quantitative approach was adopted using an online survey. Participants received a clear explanatory text about the experimental process, user interface design, and operation of ChatGPT. During the survey, respondents interacted with the ChatGPT digital agent.

In Phase A, before the interaction, participants' ATAI was assessed. After the interaction, in Phase B, various factors were evaluated: expectation confirmation, security and privacy, visual characteristics, PEOU, and trust. The purpose of Phase A was to assess users' attitudes toward AI in preparation for evaluating the research hypotheses in Phase B. At this stage, the ATAI questionnaire was administered to all participants.

In Phase B, the experiment was conducted using ChatGPT. Each participant received a list of questions to use in communication with the agent across three organizational simulation scenarios: IT support, advanced Excel functions tutorial, and personal assistant tasks. Usage duration was measured by the number of questions asked, with each participant experiencing all three scenarios. Participants could select questions from a provided list or pose their own during the simulation. They reported the number of questions asked to assess different exposure durations.

A serious agent (ChatGPT) was selected by default. User perceptions of ChatGPT's visual features and ease of use were assessed to determine its understandability and operability. ChatGPT's interface was characterized by a simple text input box for user queries, preferably in English.

Participants asked ChatGPT questions regarding information security and privacy, received reassurance about user privacy, and subsequently inquired about data retention.

To investigate the impact of expectation levels, participants were divided into two groups, each receiving different explanatory texts aimed at influencing the expectations of the agent (see Appendix C for the texts provided): (1) High Expectation Group: Received positive texts without highlighting trust-reducing features. (2) Realistic Expectation Group: Received balanced texts emphasizing potential trust-reducing aspects of the agent.

After the scenarios, participants completed scales related to expectation confirmation, security and privacy, visual characteristics, PEOU, and trust. This stage also included manipulation checks and examined the impact of expectation levels on the relationship between expectations and trust in the digital agent.

Measures

The following variables were collected using study-specific questionnaires, along with demographic data (gender, age, education level, and field of study):

- Independent Variable:
  - User Expectations from the Digital Agent were measured using the ECM scale by Følstad et al. (2021) on a 5-point Likert scale (0 = "Strongly Disagree" to 5 = "Strongly Agree"). Example item: "My experience with the chatbot exceeded my expectations." ($\alpha = 0.824$).
- Moderator Variables:
  - Expectation Setting: Participants were randomly divided into two groups, each receiving different preparatory texts.
  - User Attitude Toward AI: Measured using the ATAI scale by Sindermann et al. (2022) on an 11-point Likert scale. Example item: "I am afraid of artificial intelligence." ($\alpha = 0.724$).
  - Information Usage Reporting: Average responses to five privacy-related questions (Kwangsawad & Jattamart, 2022). Example item: "I think talking to a chatbot will lead to the dissemination of private information." ($\alpha = 0.825$).
  - Visual Characteristics and PEOU: Average responses to six items (Silva et al., 2023). Example item: "I find it easy to operate chatbots." ($\alpha = 0.933$).
  - Duration of Interaction: Measured by the average number of questions asked across the three scenarios.
- Dependent Variable:
  - Trust in the Digital Agent, measured using the scale by Gefen et al. (2003). Example item: "I believe the digital agent is dependable." ($\alpha = 0.868$).
- Control Variables:
  - Gender, age, education level, and field of study.

Results

Preliminary analyses

Before hypothesis testing, using SPSS AMOS Ver. 29.0, a confirmatory factor analysis (CFA) was performed to test the discriminant validity of the continuous scales. The CFA included five factors. The five-factor measurement model showed a good fit with the data (χ^2 = 315.5; $df = 202$; $\chi^2/df = 1.56$; CFI = 0.93; TLI = 0.91; IFI = 0.93; RMSEA = 0.06; PClose=0.02) (Hu & Bentler, 1999).

Following the CFA, the analysis continued with testing the overall model using the full sample ($N = 118$). The hypothesized relationships among expectations, mediators (attitude, transparency, visual design, exposure), and trust were assessed using a parallel mediation model with a moderated path. Subgroup analyses, such as comparisons between expectation-setting conditions or study fields, were treated as exploratory due to limited sample sizes and were reported with caution.

Descriptive analysis using SPSSTM

Table 3 provides a detailed overview of the descriptive statistics and intercorrelations among the key variables in the study, offering empirical support for the hypothesized relationships and highlighting significant associations that inform the subsequent analysis.

Regression and mediation analysis results

As presented in Table 4, the regression and mediation analysis results provide detailed estimates of the direct and indirect effects among the study variables, including coefficients, standard errors, confidence intervals, and significance levels. This table enables a clear assessment of which hypothesized pathways are empirically supported and highlights the mediating roles of key variables in shaping trust in digital agents. By systematically summarizing both direct and indirect effects, Table 4 offers empirical validation for the hypothesized relationships and clarifies the mechanisms through which user expectations, attitudes, and design features influence trust.

Hypothesis testing

Linear regression analysis was conducted using SPSS Macro (version 29) with the PROCESS v4.2 software by Hayes, Model 5, to test the hypotheses.

Hypothesis H1 suggested a positive correlation between user expectations of the digital agent and the level of trust in the digital agent. The results show that the relationship was positive but not significant; therefore, this hypothesis was not supported ($b = 0.15$, $t(106) = 0.65$, $p > .05$).

Hypothesis H1a suggested that expectation setting would moderate the relationship between user expectations of the digital agent and the level of trust in the digital agent. The results show that the relationship

Table 3
Descriptive statistics and intercorrelations.

Variable	Mean	SD	1	2	3	4	5	6	7	8	9	10
1. Expectations from the agent	3.59	1.21										
2. Attitude toward artificial intelligence (AI)	5.92	1.06	−0.16									
3. Degree of trust in the agent	3.44	1.21	0.42 **	0.16								
4. Visual characteristics	4.91	1.37	0.61 **	−0.05	0.50 **							
5. Extent of reporting the use of information	3.47	1.11	−0.24 **	0.15	0.19 *	−0.19 *						
6. Create a level of expectation	1.42	0.5	0.06	−0.07	0.09	−0.08	0.03					
7. Duration of exposure to the agent	4.34	1.58	0.06	0.02	0.05	0.01	0.01	−0.03				
8. Gender	1.19	0.4	0.13	−0.12	−0.01	−0.03	−0.1	−0.24	−0.12			
9. Age	2.3	0.6	−0.15	0.05	0.13	−0.06	0.2 *	0.01	−0.08	0.04		
10. Education	2.39	0.8	−0.03	0.07	0.11	−0.08	0.3 **	−0.05	−0.16	0	0.35 **	
11. Field of education	1.59	0.49	0.03	0.16	−0.01	−0.12	−0.03	−0.07	−0.04	0.15	0.04	0.19 *

$N = 118$; * $p<.05$; ** $p<.01$

Table 4
Regression and mediation analysis results.

Hypothesis/Path	Coefficient (b)	SE	T	p-value	LLCI	ULCI	Supported?
Direct Effects (Path A)							
EXPECT → ATT	−0.1222	0.0814	−1.5016	0.136	−0.2834	0.039	No
EXPECT → PRIVACY	−0.1935	0.0812	−2.3830	0.019	−0.3545	−0.0326	Yes
EXPECT → VISUAL	0.7210	0.0838	8.5997	0.000	0.5549	0.8872	Yes
EXPECT → EXPOSURE	0.0882	0.1229	0.7173	0.475	−0.1554	0.3317	No
Direct Effects (Path B)							
ATT → TRUST	0.1970	0.0889	2.2172	0.029	0.0208	0.3732	Yes
PRIVACY → TRUST	0.2999	0.0884	3.3943	0.001	0.1247	0.4751	Yes
VISUAL → TRUST	0.3518	0.0875	4.0219	0.0001	0.1784	0.5253	Yes
EXPOSURE → TRUST	0.0321	0.0582	0.5521	0.582	−0.0832	0.1474	No
EXPECT → TRUST (Direct)	0.1534	0.2352	0.6521	0.516	−0.3130	0.6197	No
Moderation Effects							
EXPECT × GROUP → TRUST	0.0874	0.1505	0.5805	0.563	−0.2110	0.3858	No
Indirect Effects (Mediation)							
EXPECT → ATT → TRUST	−0.0241	0.0238	–	–	−0.0822	0.0083	No
EXPECT → PRIVACY → TRUST	−0.0581	0.0338	–	–	−0.1360	−0.0046	Yes
EXPECT → VISUAL → TRUST	0.2537	0.0768	–	–	0.1056	0.4070	Yes
EXPECT → EXPOSURE → TRUST	0.0028	0.0107	–	–	−0.0177	0.0281	No

was positive but not significant; therefore, this hypothesis was not supported ($b = 0.09$, t(106) = 0.58, $p > .05$).

Hypothesis H2a suggested a positive correlation between users' attitudes toward AI and their expectations of the digital agent. The results show that the relationship was negative and not significant; therefore, this hypothesis was not supported ($b = -0.12$, t(106) = −1.5, $p > .05$).

Hypothesis H2b suggested a positive correlation between users' attitudes toward AI and trust in the digital agent. The results show that the relationship was positive and significant; therefore, this hypothesis was supported ($b = 0.19$, t(106) = 3.00, $p < .05$).

Hypothesis H2c suggested that users' attitudes toward AI would mediate the relationship between their expectations of the digital agent and the level of trust in the agent. The mediation effect was not confirmed; therefore, this hypothesis was not supported (Indirect effect = −0.024, SE = 0.02, 95 % CI [−0.082, 0.008]).

Hypothesis H3a suggested a positive correlation between the degree of information disclosure and users' expectations of the digital agent. The results show that the relationship was negative and significant; therefore, this hypothesis was not supported ($b = -0.19$, t(112) = −2.4, $p < .05$).

Hypothesis H3b suggested a positive correlation between the degree of information disclosure and the level of trust in the digital agent. The results show that the relationship was positive and significant; therefore, this hypothesis was supported ($b = 0.30$, t(106) = 3.39, $p < .01$).

Hypothesis H3c suggested that the degree of information disclosure mediates the relationship between user expectations of the digital agent and the level of trust in the agent. The mediation effect was confirmed; therefore, this hypothesis was supported (Indirect effect = −0.06, SE = 0.034, 95 % CI [−0.136, −0.005]).

Hypothesis H4a suggested a positive correlation between visual features and PEOU and users' expectations of the digital agent. The results show that the relationship was positive and significant; therefore, this hypothesis was supported ($b = 0.72$, t(112) = 8.6, $p < .01$).

Hypothesis H4b suggested a positive correlation between visual features and PEOU and the level of trust in the digital agent. The results indicate a positive and significant relationship; therefore, this hypothesis was supported ($b = 0.35$, t(106) = 4.02, $p < .01$).

Hypothesis H4c suggested that visual features and PEOU mediate the relationship between user expectations of the digital agent and the level of trust in the agent. The mediation effect was confirmed; therefore, this hypothesis was supported (Indirect effect = 0.25, SE = 0.08, 95 % CI [.106, 0.407]).

Hypothesis H5a suggested a positive correlation between exposure duration to the agent and users' expectations of the digital agent. The results show that the relationship was positive but not significant; therefore, this hypothesis was not supported ($b = 0.09$, t(112) = 0.72, $p > .05$).

Hypothesis H5b suggested a positive correlation between the duration of exposure to the agent and the degree of trust in the digital agent. The results show that the relationship was positive but not significant; therefore, this hypothesis was not supported ($b = 0.03$, t(110) = 0.55, $p > .05$).

Hypothesis H5c suggested that the duration of exposure to the agent mediates the relationship between user expectations of the digital agent and the degree of trust in the digital agent. The mediation effect was not confirmed; therefore, this hypothesis was not supported (Indirect effect = 0, SE = 0.01, 95 % CI [−0.018, 0.028]).

Control Variables (Covariates)

All of the control variables—that is, gender, age, education level, and field of education—were found to be insignificant.

Subgroup analyses

No statistically significant differences were found between the high-expectation and realistic-expectation groups, nor between participants from technical and non-technical backgrounds.

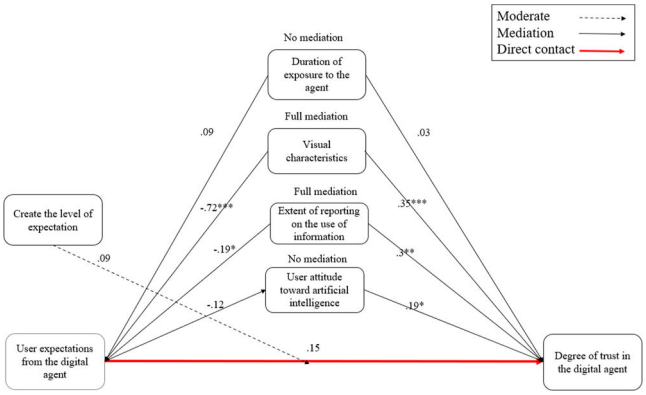


Fig. 2. Hypothesis testing results.
(***Significant at the 0.001 level; **Significant at the 0.01 level; *Significant at the 0.05 level).

Summary of findings

The results demonstrate that while user expectations do not directly correlate with trust in digital agents like ChatGPT within organizational contexts, factors such as transparency regarding information usage and effective visual design play crucial roles in fostering trust among users. In particular, these factors (information usage, transparency, and visual design) also mediate the relationship between user expectations and trust in the digital agent. Additionally, the study found a significant positive correlation between user attitudes toward AI and trust in the digital agent. These findings underscore the complex dynamics involved in building user trust in AI-powered digital agents and highlight the importance of considering multiple factors in their design and implementation.

Fig. 2 below shows the results of the relationships in the model:

As illustrated in Fig. 2, the research model reveals both direct and mediated pathways through which user expectations influence trust in the digital agent. Visual characteristics, information reporting, and user attitudes serve as key mediators, with visual design showing the strongest significant mediation effect. Additionally, the model identifies a moderating role of expectation creation in shaping the trust relationship.

Discussion

This study investigates the factors influencing trust in AI-driven digital agents, focusing on ChatGPT within organizational contexts. The findings highlight that transparency in information usage and effective visual design are critical determinants of trust. Transparency not only directly fosters trust but also mediates the relationship between user expectations and trust, underscoring the importance of clear communication about data usage and privacy policies. Similarly, intuitive and visually coherent interfaces enhance user experience and serve as mediators in trust-building, aligning with the TAM's emphasis on PEOU.

Positive attitudes toward AI are strongly correlated with higher trust levels, supporting the ATAI framework, which posits that favorable perceptions of AI drive user engagement and adoption. However, contrary to traditional models such as ECT, user expectations alone do not directly predict trust in ChatGPT. This challenges ECT, which posits that meeting or exceeding expectations leads to satisfaction and trust (Bhattacherjee, 2001). The findings suggest that while expectations are inevitable, they may be overshadowed by other factors, such as transparency and usability, in shaping trust.

Although conventional models such as ECT suggest that meeting or exceeding user expectations should foster trust, the findings challenge this assumption. In this study, expectations alone do not significantly

predict trust in AI-enhanced digital agents. This suggests that, in the context of AI systems, factors like transparency and usability may outweigh expectation alignment in shaping user trust. Recent literature emphasizes that users' trust decisions are increasingly influenced by how the system communicates its logic and how intuitive it feels to use, rather than by whether it simply meets expectations (Jacovi et al., 2021; Lockey et al., 2021). These findings imply a theoretical shift from outcome-based trust models to design-centered frameworks that prioritize user empowerment and transparency (Guidotti et al., 2018; Koene et al., 2019).

Future research should employ immersive techniques—such as scenario-based role-play, repeated interventions, or longitudinal designs—to better capture the dynamics of trust in real-world organizational settings. These approaches may help overcome the limitations of brief expectation manipulations, especially as user attitudes toward AI are increasingly shaped by prior exposure and media narratives. This direction aligns with emerging calls for updated theoretical frameworks that prioritize actionable transparency and intuitive usability over expectation management alone, in line with evolving governance standards such as the EU AI Act (Koene et al., 2019; Lockey et al., 2021).

The study also found that exposure duration to the digital agent does not significantly impact trust formation, challenging the assumption that repeated interactions inherently build trust. Instead, the quality of interaction appears to be more critical than the quantity. The ineffectiveness of brief expectation manipulations may be explained by the influence of prior exposure to AI and prevailing media narratives, which shape user expectations more powerfully than experimental interventions (Kosch et al., 2023).

Several contextual factors may limit the generalizability of these findings. The sample was drawn from Israeli organizations, which are characterized by a high-tech orientation and early adoption culture, potentially differing from more risk-averse environments. Attitudes toward AI in Israeli workplaces may reflect a unique cultural perspective on technology adoption (Lartey, 2021). The predominantly male sample may also have influenced the results, as men are generally more receptive to technological innovations and exhibit greater risk tolerance (Nguyen et al., 2021).

Furthermore, the study focuses exclusively on ChatGPT, and its cross-sectional design provides only a snapshot of trust formation at a single point in time. Trust in AI agents is likely dynamic, evolving with ongoing interaction, which underscores the need for longitudinal research.

Participant feedback reflects both enthusiasm and uncertainty when interacting with ChatGPT, suggesting that future studies could benefit from qualitative methods to better capture users' emotional and cognitive responses. Theoretically, the study advances understanding by integrating ECT and TAM, demonstrating that while expectation alignment is necessary, usability and transparency are central to trust-building in AI. The findings reinforce the need for organizations to prioritize transparent communication of data policies, compliance with privacy regulations such as the GDPR (Mhlanga, 2023), and design intuitive user interfaces that align with user needs to enhance PEOU and build confidence in digital agents (Silva et al., 2023). Educational initiatives to improve public attitudes toward AI can further promote adoption by addressing fears and misconceptions (Sindermann et al., 2022). Rather than relying solely on repeated exposure, organizations should focus on delivering high-quality interactions that meet user needs efficiently.

This study provides valuable insights into the determinants of trust in AI-enhanced digital agents such as ChatGPT within organizational contexts. By highlighting the roles of transparency, visual design, and user attitudes toward AI, it provides actionable recommendations for enhancing user engagement and promoting the adoption of AI technologies. These findings are particularly relevant for addressing key challenges related to privacy concerns, usability barriers, and the design of intuitive, trustworthy AI systems.

Beyond academic contributions, the study results provide actionable guidance for both AI developers and policymakers. Developers should prioritize transparent design and user-centered visual interfaces to foster trust, ensuring that users understand how AI systems operate and how their data are handled (Lockey et al., 2021; Silva et al., 2023). Incorporating explainability features and clear communication about system logic and limitations is essential for trustworthy AI (Grimes et al., 2021; Shin, 2021). User-centered design—including intuitive, visually coherent interfaces—empowers users and reduces uncertainty, which in turn enhances trust and adoption (Silva et al., 2023).

For developers and policymakers, the results emphasize the importance of transparent design, explainability, and user-centered interfaces to foster trust. Policymakers should support robust governance frameworks, including disclosure norms and explainability standards, as reflected in recent regulations such as the EU AI Act (Lockey et al., 2021; Nemitz, 2018). Transparency and accountability are foundational for public trust and should be tailored to the needs of different stakeholders (Felzmann et al., 2019). Public education and engagement are also essential for overcoming misconceptions and fostering informed trust in AI systems (Dignum, 2019).

Future research should involve cross-cultural and longitudinal studies to explore how trust in AI agents develops and changes over time. Additional factors such as emotional intelligence, personalization, and ethical transparency may provide deeper insights into the mechanisms shaping user trust in AI-powered tools.

Conclusion

This study elucidates the determinants of trust in AI-driven digital agents such as ChatGPT within organizational contexts, drawing on the ECT, TAM, and the ATAI framework. The findings demonstrate that trust is shaped by a dynamic interplay of transparency, usability, and general attitudes toward AI rather than by user expectations or exposure duration alone. Transparency in information usage and effective visual design emerged as pivotal factors, mediating the relationship between user expectations and trust and highlighting the importance of aligning user perceptions with system performance. Positive attitudes toward AI are also strongly correlated with higher trust levels, underscoring the need to address negative perceptions and misconceptions about AI technologies.

From a practical perspective, the study suggests that organizations seeking to enhance user engagement and adoption of AI technologies should prioritize transparent communication about data usage and ensure compliance with privacy regulations such as the GDPR. Designing intuitive interfaces that align with user needs can enhance PEOU and build confidence in digital agents. Educational initiatives aimed at improving public attitudes toward AI may further mitigate fears and misconceptions, fostering greater acceptance and satisfaction.

Despite its contributions, the study has several limitations. The sample predominantly comprises participants from Israeli organizations, which may limit the generalizability of the findings to other cultural or organizational contexts. Using simulated workplace scenarios, while offering control and replicability, may not fully capture the complexities of real-world interactions with AI agents such as ChatGPT. Additionally, the relatively brief exposure to the digital agent may not reflect the longer-term processes involved in trust formation and sustained use. The modest sample size and its demographic characteristics, particularly the underrepresentation of female participants, may have influenced some outcomes, such as attitudes toward AI and reported levels of trust. These factors are especially relevant for subgroup analyses, which should be interpreted with caution.

Future research should aim to replicate these findings across larger and more diverse samples within varied cultural and professional environments. Longitudinal and mixed-method approaches are recommended to deepen understanding of trust dynamics and capture changes over time in organizational human–AI collaboration. Further

investigation into factors such as emotional intelligence, personalization features, and ethical considerations could also provide deeper insights into user trust in AI-driven digital agents.

In conclusion, this study offers valuable insights into the determinants of trust in AI-driven digital agents like ChatGPT within organizational settings. By highlighting the roles of transparency, visual design, and attitudes toward AI, it provides actionable recommendations for organizations aiming to enhance adoption and user satisfaction. Ultimately, fostering trust in digital agents is essential for their successful integration into workflows and for realizing their potential to improve work performance and user experiences.

Limitations and future research

Despite its contributions, this study has limitations. The Israeli sample, relevant for tech-forward environments, may not generalize across cultures. The single-country and male-dominant sample may skew perceptions of trust and AI. The manipulation of expectations had limited effect, possibly due to weak framing or participants' prior exposure to AI. Future research should test stronger framing interventions, adopt longitudinal designs, and include qualitative methods (e.g., interviews or focus groups) to gain richer insights into trust dynamics.

To complement the survey-based findings and address the complexity of trust formation in AI-driven digital agents, future research should incorporate qualitative methods, such as in-depth interviews or focus groups. These approaches can provide deeper insights into users' emotional responses, motivations, and contextual factors influencing trust, which are often not fully captured by quantitative measures.

Data availability statement

Data are available upon request from the authors.

CRediT authorship contribution statement

Iris Glassberg: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Yael Brender-Ilan:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Moti Zwilling:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation.

Declaration of competing interest

No potential conflict of interest was reported by the author.

Appendices

Appendix A. – Questionnaires

Følstad's ECM Expectations Confirmation Model Scale (2021).
CON1 My experience with this chatbot was greater than my expectations.
CON2 The service level provided by this chatbot was greater than what I expected.
CON3 In general, most of my expectations from using this chatbot were confirmed.
Scale of Users' Attitudes Toward Artificial Intelligence – Sindermann's ATAI Scale (2022).
AI01 I fear artificial intelligence (AI).
AI02 I trust AI.
AI03 AI will destroy humankind.
AI04 AI will benefit humankind.
AI05 AI will cause many job losses.
Extent of Information Use – Privacy Scale (Kwangsawad & Jattamart, 2022)

PPR1 I think chatbot conversations can disclose personal information that is to be published.
PPR2 I think chatbot conversations may have collected my personal information.
PPR3 I recognize that disclosing personal information through chatbots is a risk.
PPR4 I recognize that disclosing personal information through chatbots can have a negative impact on me.
PPR5 Privacy risks are an essential part of my next chatbot decision.

Visual Characteristics and PEOU Scale – Silva et al. (2023).
PEU1: My interaction with chatbots is clear and understandable.
PEU2: Interacting with chatbots requires less mental effort.
PEU3: I find chatbots easy to use.
PEU4: I find it easy to get the chatbots to do what I want them to do.
PEU5: It is easy for me to become adept at using chatbots.
PEU6: I possess the knowledge necessary to use chatbots.
Degree of Trust in Digital Agent – Gefen's Trust Scale (2003).
TRU1 I believe that this chatbot is trustworthy.
TRU2 I do not doubt the honesty of the information provided by this chatbot.
TRU3 I feel assured that this chatbot service can protect users.
TRU4 Overall, I have trust in this chatbot.

Appendix B. – Operating scenarios

Chatbot activation scenario for IT support:
Please select one or more of the questions and ask ChatGPT – write "Simulate IT support chatbot" and ask in Hebrew:

1. How do I reset my email account password?
2. I forgot my password; how do I reset it?
3. How do I connect to the Wi-Fi network?
4. Can you recommend a good antivirus for your computer?
5. How do I transfer files from one computer to another?
6. My computer is running slowly, what should I do?
7. How do I update my operating system?
8. Is there a way to recover a deleted file from my computer?
9. How do I back up my data?
10. How do I connect my printer to my computer?

Chatbot activation scenario for training in Excel software:
Please select one or more of the questions and ask ChatGPT – write "Simulate Excel Training Chatbot" and ask in Hebrew:

1. How do I create a pivot table in Excel?
2. Can you show me how to use the VLOOKUP function in Excel?
3. What is conditional formatting, and how can I use it to highlight cells in Excel?
4. How do I chart in Excel?
5. Can you explain how to use the SUMIF function in Excel?
6. What is the difference between a workbook and a worksheet in Excel?
7. How do I password-protect an Excel file?
8. Can you show me how to use the CONCATENATE function in Excel?
9. How can I use Excel to calculate a weighted average?
10. What is the difference between a relative cell reference and a fixed cell address in Excel?

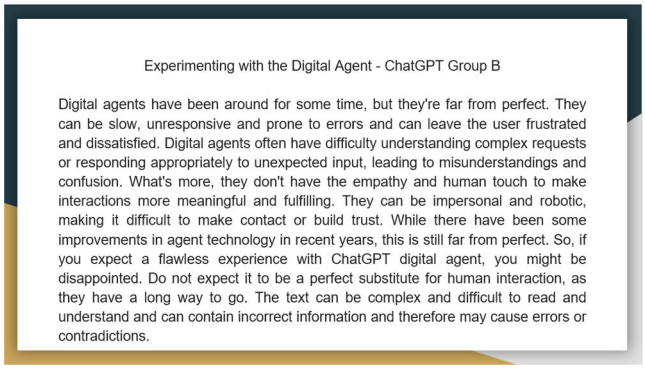
Chatbot Digital Personal Assistant activation scenario:
Please select one or more of the questions and ask ChatGPT – write "Simulate Virtual Personal Assistant chatbot" and ask in Hebrew:

1. Can you make an appointment with Jonathan for next Wednesday at 2 pm.?
2. What is the climate in Tel Aviv today?
3. Can you remind me to call my dentist tomorrow at 10 a.m.?
4. What are the exchange rates between the dollar and the shekel?
5. Can you order me a pizza from my favorite pizzeria?
6. Can you give me directions to the nearest gas station?
7. What is the latest news about the COVID-19 pandemic?
8. Can you play me relaxing music?
9. What is the name of the CEO of Apple Company?
10. Can you set a timer for me for 20 min?

Appendix C. – Explanatory text about digital agents

Group A and Group B

Group A. Text to create high expectations of agent performance.



Group B. Text to create a realistic and adjusted expectation of the agent's performance.

References

Afroogh, S., Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2024). Trust in AI: Progress, challenges, and future directions. *Humanities and Social Sciences Communications, 11*(1), 1–30. https://doi.org/10.1057/s41599-025-04374-1

Ahmad, R., Siemon, D., Gnewuch, U., & Robra-Bissantz, S. (2022, January). A framework of personality cues for conversational agents. In *Proceedings of the 55th Hawaii international conference on system sciences*. https://doi.org/10.24251/HICSS.2022.524

Ahonen, E. (2020). The effects of transnationality on trust in customer service chatbots. https://aaltodoc.aalto.fi/items/82ec1e6b-2860-4b5a-b8c3-6f325f0fc76d.

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly*, 351–370. https://doi.org/10.2307/3250921

Blazevic, V., & Sidaoui, K. (2022). The TRISEC framework for optimizing conversational agent design across search, experience and credence service contexts. *Journal of Service Management*. https://doi.org/10.1108/JOSM-10-2021-0402

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319–340. https://doi.org/10.2307/249008

Diederich, S., Brendel, A. B., Morana, S., & Kolbe, L. (2022). On the design of and interaction with conversational agents: An organizing and assessing review of human-computer interaction research. *Journal of the Association for Information Systems, 23*(1), 96–138. https://doi.org/10.17705/1jais.00724

Følstad, A., Araujo, T., Law, E. L. C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., & Luger, E. (2021). Future directions for chatbot research: An interdisciplinary research agenda. *Computing, 103*(12), 2915–2942. https://link.springer.com/article/10.1007%2Fs00607-021-01016-7.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Inexperience and experience with online stores: The importance of TAM and trust. *IEEE Transactions on Engineering Management, 50*(3), 307–321. https://doi.org/10.1109/TEM.2003.817277

Gkinko, L., & Elbanna, A. (2022). The appropriation of conversational AI in the workplace: A taxonomy of AI chatbot users. *International Journal of Information Management, 69*, Article 102568. https://doi.org/10.1016/j.ijinfomgt.2022.102568

Gkinko, L., & Elbanna, A. (2023). Designing trust: The formation of employees' trust in conversational AI in the digital workplace. *Journal of Business Research, 158*, Article 113707. https://doi.org/10.1016/j.jbusres.2023.113707

Grimes, G. M., Schuetzler, R. M., & Giboney, J. S. (2021). Mental models and expectation violations in conversational AI interactions. *Decision Support Systems, 144*, Article 113515. https://doi.org/10.1016/j.dss.2021.113515

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR), 51*(5), 1–42. https://doi.org/10.1145/3236009

Gulati, S., Sousa, S., & Lamas, D. (2019). Design, development and evaluation of a human-computer trust scale. *Behaviour & Information Technology, 38*(10), 1004–1015. https://doi.org/10.1080/0144929X.2019.1656779

Hinsen, S., Hofmann, P., Jöhnk, J., & Urbach, N. (2022). *How can organizations design purposeful human-AI interactions: A practical perspective from existing use cases and interviews*. https://doi.org/10.24251/HICSS.2022.024

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021, March). Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 624–635). https://doi.org/10.1145/3442188.3445923

Jiang, Y., Yang, X., & Zheng, T. (2023). Make chatbots more adaptive: Dual pathways linking human-like cues and tailored response to trust in interactions with chatbots. *Computers in Human Behavior, 138*, Article 107485. https://doi.org/10.1016/j.chb.2022.107485

Kaya, F., Aydin, F., Schepman, A., Rodway, P., Yetişensoy, O., & Demir Kaya, M. (2022). The roles of personality traits, AI anxiety, and demographic factors in attitudes toward artificial intelligence. *International Journal of Human–Computer Interaction*, 1–18. https://doi.org/10.1080/10447318.2022.2151730

Koene, A., Clifton, C. W., Hatada, Y., Webb, H., Patel, M., Machado, C., & Reisman, D. (2019). A governance framework for algorithmic accountability and Transparency: Study. *European Parliament*. https://doi.org/10.2861/59990

Kosch, T., Welsch, R., Chuang, L., & Schmidt, A. (2023). The placebo effect of artificial intelligence in human–computer interaction. *ACM Transactions on Computer-Human Interaction, 29*(6), 1–32. https://doi.org/10.1145/3529225

Kwangsawad, A., & Jattamart, A. (2022). Overcoming customer innovation resistance to the sustainable adoption of chatbot services: A community-enterprise perspective in Thailand. *Journal of Innovation & Knowledge, 7*(3), Article 100211. https://doi.org/10.1016/j.jik.2022.100211

Letheren, K., Russell-Bennett, R., & Whittaker, L. (2020). Black, white or grey magic? Our future with artificial intelligence. *Journal of Marketing Management, 36*(3–4), 216–232. https://doi.org/10.1080/0267257X.2019.1706306

Lockey, S., Gillespie, N., Holm, D., & Someh, I. A. (2021). *A review of trust in artificial intelligence: Challenges, vulnerabilities and future directions.doi.* https://doi.org/10.24251/HICSS.2021.664

Mohd Rahim, N. I., A. Iahad, N., Yusof, A. F., & A. Al-Sharafi, M. (2022). AI-based chatbots adoption model for higher-education institutions: A hybrid PLS-SEM-neural network modelling approach. *Sustainability, 14*(19), Article 12726. https://doi.org/10.3390/su141912726

Nordheim, C. B. (2018). Trust in chatbots for customer service–findings from a questionnaire study (master's thesis). *Google Scholar*.

Nguyen, D. M., Chiu, Y. T. H., & Le, H. D. (2021). Determinants of continuance intention towards banks' chatbot services in Vietnam: A necessity for sustainable development. *Sustainability, 13*(14), 7625. https://doi.org/10.3390/su13147625

Nguyen, T. (2019). Potential effects of chatbot technology on customer support: A case study. *Google scholar*.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of marketing research, 17*(4), 460–469. https://doi.org/10.1177/002224378001700405

Pinto, A., Sousa, S., Simões, A., & Santos, J. (2022). A trust scale for Human-robot interaction: Translation, adaptation, and validation of a Human computer trust scale. *Human Behavior and Emerging Technologies, 2022*. https://doi.org/10.1155/2022/6437441

Pesonen, J. A. (2021, July). 'Are you OK?'Students' trust in a chatbot providing support opportunities. In *Learning and collaboration technologies: Games and virtual environments for learning: 8th international conference, LCT 2021, held as part of the 23rd HCI international conference, HCII 2021, virtual event, july 24–29, 2021, proceedings, part II* (pp. 199–215). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-77943-6_13.

Riethorst, J. A. (2022). The influence of perceived similarity of avatars on trust in medical chatbots. *Google Scholar*.

Rathore, B. (2023). Future of AI & Generation Alpha: ChatGPT beyond boundaries. *Eduzone: International Peer Reviewed/Refereed Multidisciplinary Journal, 12*(1), 63–68. https://doi.org/10.56614/eiprmj.v12i1y23.254

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies, 146*, Article 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., & Montag, C. (2021). Assessing the attitude towards artificial intelligence: Introduction of a short measure in german, chinese, and english language. *KI-Künstliche intelligenz, 35*, 109–118. https://doi.org/10.1007/s13218-020-00689-0

Sindermann, C., Yang, H., Elhai, J. D., Yang, S., Quan, L., Li, M., & Montag, C. (2022). Acceptance and fear of artificial intelligence: Associations with personality in a German and a Chinese sample. *Discover Psychology, 2*(1), 8. https://doi.org/10.1007/s44202-022-00020-y

Silva, F. A., Shojaei, A. S., & Barbosa, B. (2023). Chatbot-based services: A study on customers' Reuse intention. *Journal of Theoretical and Applied Electronic Commerce Research, 18*(1), 457–474. https://doi.org/10.3390/jtaer18010024

Terblanche, N., & Kidd, M. (2022). Adoption factors and moderating effects of age and gender that influence the intention to use a non-directive reflective coaching chatbot. *SAGE Open, 12*(2), Article 21582440221096136. https://doi.org/10.1177/21582440221096136