Rapid #: -23789974

CROSS REF ID: **6066101**

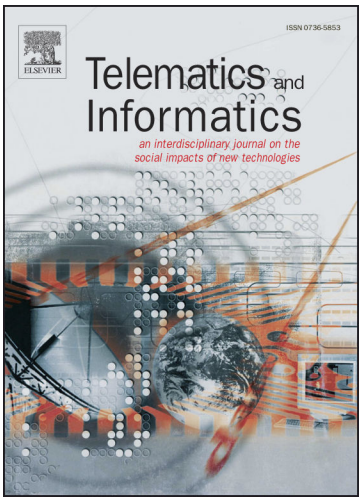
LENDER: **GZN (University of Wisconsin - Milwaukee) :: Ejournals**

BORROWER: **LUU (Louisiana State University) :: Main Library**

TYPE:	Article CC:CCG
JOURNAL TITLE:	Telematics and informatics
USER JOURNAL TITLE:	Telematics and Informatics
ARTICLE TITLE:	Trust in AI-driven chatbots: A systematic review
ARTICLE AUTHOR:	Ting Ng, Sheryl Wei,
VOLUME:	Pre-print
ISSUE:	
MONTH:	
YEAR:	2025
PAGES:	102240-
ISSN:	0736-5853
OCLC #:	

Processed by RapidX: 1/13/2025 1:02:18 PM

Journal Pre-proofs



Trust in AI-driven chatbots: A systematic review

Sheryl Wei Ting Ng, Renwen Zhang

PII: S0736-5853(25)00002-4
DOI: https://doi.org/10.1016/j.tele.2025.102240
Reference: TELE 102240

To appear in: *Telematics and Informatics*

Received Date: 3 August 2024
Revised Date: 18 November 2024
Accepted Date: 6 January 2025

Please cite this article as: Ting Ng, S.W., Zhang, R., Trust in AI-driven chatbots: A systematic review, *Telematics and Informatics* (2025), doi: https://doi.org/10.1016/j.tele.2025.102240

Trust in AI-driven chatbots: A systematic review

Abstract

Advancements in artificial intelligence (AI) have enabled increasingly natural and human-like interactions with chatbots and conversational agents. However, the processes and mechanisms underlying trust development in AI chatbots remain underexplored. This study provides a systematic review of how trust in AI-driven chatbots is defined, operationalised, and studied, synthesizing factors influencing trust development and its outcomes. An analysis of 40 articles revealed notable variations and inconsistencies in trust conceptualisations and operationalisations. Predictors of trust are categorized into five groups: user, machine, interaction, social, and context-related factors. Trust in AI-driven chatbots leads to diverse outcomes that span affective, relational, behavioural, cognitive, and psychological domains. The review underscores the need for longitudinal studies to better understand the dynamics and boundary conditions of trust development. These findings offer valuable insights for advancing human-machine communication (HMC) research and informing the design of trustworthy AI systems.

Keywords: Trust, Chatbots, Conversational agents, Artificial intelligence, Human-machine communication, Systematic review

1. Introduction

Artificial agents capable of engaging in natural language conversations with humans are commonly known as chatbots, conversational agents, or interactive agents. The advancement of artificial intelligence (AI) has made interactions with chatbots increasingly human-like (Chaturvedi et al., 2023). Trust plays a critical role in the study of human-machine communication (HMC), as it is essential for establishing effective and reliable interactions between humans and machines (Yang & Wibowo, 2022). Trust also affects various aspects of user-technology interactions and is essential for adoption and continued usage of new technologies (Yang & Wibowo, 2022). Therefore, understanding how trust is developed is vital to ensure optimal outcomes.

Moreover, as AI-driven technologies rapidly evolve, understanding the outcomes of trust in these technologies becomes increasingly important. Studies have shown that excessive trust can lead to overreliance on these machines (Asan et al., 2020; Prahl & Van, 2021). Ethical concerns such as user privacy, deception, and the absence of an ethical framework to regulate AI are critical issues raised by researchers focusing on ethical AI (Denecke et al., 2021). These concerns highlight potential negative consequences that may still be undiscovered. Therefore, while building trust in machines is beneficial, it is equally important to mitigate the risks associated with excessive trust in these technologies.

However, the theoretical foundation supporting users' social reactions towards these machines remains uncertain. Specifically, there is a lack of consensus on the concept and measurement of trust in HMC. Existing work suggests that the foundational concepts of trust derived from human relationships are insufficient for understanding trust in HMC because of ontological differences between humans and machines (Guzman, 2020; Madhavan & Wiegmann, 2007). Furthermore, current reviews that attempt to consolidate various perspectives and definitions of trust adopt a broad view of AI machines, with reviews specifically focused on trust in AI-driven chatbots being scarce, and the implications of trust being overlooked. This is concerning given that trusting AI-driven chatbots differs significantly from trusting other types of AI systems in terms of credibility attribution (Pentina et al., 2023), attachment formation (Li & Zhang, 2024), and anthropomorphic elements (Chen & Park, 2021). Therefore, this systematic review seeks to understand how trust in AI-driven chatbots is defined, operationalised, and studied, and to synthesize the factors that influence trust development and the outcomes of trust in AI-driven chatbots. This study has significance for advancing the understanding of trust in AI-driven chatbots and has implications for future research that seeks to advance HMC research and practice.

2. Literature Review

2.1 The foundational concept of trust

Before diving into trust in HMC, it is necessary to first provide an overview of the foundational concepts that have influenced the development of trust in AI machines, including interpersonal trust, institutional trust, and swift trust. Interpersonal trust forms the bedrock of human relationships, characterized by dimensions such as competence, benevolence, and integrity (Mayer et al., 1995). It is typically developed over time through repeated interactions and is essential for fostering cooperation and social cohesion (Rotter, 1967). This form of trust is dynamic and deeply personal, evolving as individuals share experiences and build a history of reliable interactions. Institutional trust goes beyond individual relationships to include trust in larger systems, such as governments,

organizations, and public institutions. This form of trust is essential for the operation of intricate societies, where individuals depend on institutions to handle risks and offer services (Giddens, 1990). Unlike interpersonal trust, institutional trust is often based on the perceived legitimacy and effectiveness of these institutions. It relies on the belief that these larger entities operate in a trustworthy manner, even without personal interaction with the individuals within them (Giddens, 1990). Swift trust is a concept that applies to temporary, goal-oriented teams or contexts where trust must be established quickly to achieve specific objectives (Meyerson et al., 1996). Unlike interpersonal trust, which evolves slowly and is deeply rooted in personal experience, swift trust is built quickly and often based on initial impressions, roles, and expectations. It is characterized by a provisional and tentative nature, relying heavily on the immediate performance and perceived competence of team members or collaborators in the initial stages of interaction (Meyerson et al., 1996).

These foundational concepts of trust each provide valuable insights into understanding trust in AI. Interpersonal trust parallels the trust users might develop in AI systems based on their perceived competence, benevolence, and integrity. For example, an AI-driven chatbot that demonstrates reliable performance (competence), acts in the user's best interest (benevolence), and adheres to transparent and ethical guidelines (integrity) is more likely to be trusted by users (Mayer et al., 1995). Extending this framework, predictability has emerged as a critical dimension in trust dynamics (ABI+), reflecting users' expectations for consistent and reliable behaviour from AI systems (Dietz et al., 2006; Toreini et al., 2020). This additional element underscores how trust is sustained through stable and expected performance over time.

Institutional trust extends this by highlighting the importance of the broader context in which AI operates. Trust in AI technologies is not solely based on the AI's performance but also on the trustworthiness of the institutions that create and manage these technologies (Giddens, 1990). Users need assurance that these institutions uphold high ethical standards, protect user data, and operate transparently before accepting and adopting these AI technologies. Swift trust provides a framework for understanding how trust in AI can be established quickly in specific contexts. For example, when users interact with a new AI system, they may initially trust it based on its immediate performance and the context in which it is used (Meyerson et al., 1996). If an AI-driven chatbot provides quick, accurate, and helpful responses in its initial interactions, users are likely to develop a provisional trust that can either be strengthened or weakened by subsequent interactions.

However, there are also ontological differences between humans and machines that suggest that trust in HMC is unique from other forms of trust discussed earlier. According to Guzman (2020), the human-computer divide lies in areas like the origin of being, autonomy, and emotions, all of which can affect interactions with machines. Madhavan and Wiegmann (2007) showed that trust in human-human relationships develops in different sequential manners from human-automation relationships, and this is echoed by Lee and See (2004). For example, the primary basis of trust in automation involves judgements of performance, but interpersonal trust is based on the dispositional characteristics of the human advisor (Madhavan & Wiegmann, 2007). Furthermore, interpersonal trust often arises within social exchange relationships, but this is a dynamic that lacks symmetry in the context of trust between humans and machines because machines do not have intentions or motivations in the same way that humans do (Lee & See, 2004). Advancements in technology further explain why trust in HMC should differ from institutional trust. AI-driven chatbots capable of unscripted free speech have more autonomy. While this may amplify the anthropomorphism

people attribute to machines and increase feelings of interpersonal trust, it also minimizes the role of the creators of these systems, and consequently, reduces the focus on institutional trust.

Moreover, as AI machines continue to evolve in their roles as social companions (Chaturvedi et al., 2023), user interactions with them are expanding beyond just a single touch point or to achieve a specific goal, which renders swift trust less applicable. As such, while theories from human-human communication may be useful starting points in HMC research, they may not be entirely replicable to trust in HMC. Consequently, this dual recognition of both the similarities and the ontological differences necessitates a more nuanced approach to studying trust in AI. This includes acknowledging that while AI systems can mirror human trust dynamics, they also introduce unique elements that must be distinctly understood.

2.2 Existing reviews of trust in AI technologies and AI-driven chatbots

Trust in AI technologies is a complex, multifaceted concept, which has prompted several attempts to consolidate different perspectives and definitions of trust. Existing systematic reviews, such as those conducted by Bach et al. (2022), have explored user trust across a broad range of AI technologies, identifying diverse definitions and influential factors like socio-ethical considerations, technical and design features, and user characteristics. Furthermore, Yang and Wibowo (2022) provided an overview of the components of trust and the critical factors that influence users' trust in AI, but adopted a broad definition of AI that encompassed many applications. Likewise adopting a broad perspective of AI machines, Ueno et al. (2022) conducted a scoping review and found that studies generally lacked an explicit definition of trust. Ueno et al. (2022) also revealed the wide variety of trust measures used across studies. Problems with trust measurements have been identified even in other domains of research (Brzowski & Nathan-Roberts, 2019; Zhang et al., 2023). There is a lack of attention on the outcomes of trust (Bach et al., 2022; Ueno et al., 2022). Among existing reviews, only Yang and Wibowo (2022) identified outcomes of trust in AI relating to cognitive, affective, and behavioural changes.

However, reviews focusing specifically on trust in AI-driven chatbots are relatively scarce. Existing reviews about AI-driven chatbots either completely neglect trust (Kuhail et al., 2022; Okonkwo & Ade-Ibijola, 202) or only mention it briefly, rather than exploring the concept in depth (Jenneboer et al., 2022; Rapp et al., 2021). One notable exception is Rheu et al.'s (2021) review, which shed light on the conversational agents' attributes and users' characteristics that enhance trust, including the agents' social intelligence and communication style. While their findings are valuable for informing the factors that contribute to trust formation, they neglect to understand the varied conceptualisations and operationalisations of trust and the implications of trusting these chatbots.

Trusting AI-driven chatbots differs significantly from trusting other types of AI systems. First, unlike recommendation engines or prediction tools, AI-driven chatbots engage users in dynamic and human-like conversations, which often leads to higher credibility attribution (Pentina et al., 2023; Sundar, 2008). This interactivity fosters interpersonal dynamics, where users may form bonds with chatbots (Xie & Pentina, 2022; Li & Zhang, 2024). In contrast, trust in standard AI systems typically revolves around their functional performance rather than social or emotional connections (Yang & Wibowo, 2022). Second, AI-driven chatbots frequently incorporate strong anthropomorphic elements, displaying more human-like conversational styles across diverse topics, unlike the rigid, pre-defined

conversations in rule-based AI systems (Jo et al., 2023). Studies suggest that anthropomorphic agents are perceived as more trustworthy (Chen & Park, 2021), indicating that AI-driven chatbots are particularly effective at building trust among users.

Thus, AI-driven chatbots present significant risks due to their intelligence and ubiquity. Their ability to generate realistic and coherent responses can be exploited for malicious purposes, such as spreading misinformation (Weidinger et al., 2021). These chatbots may also perpetuate the biases inherent in their training data (Weidinger et al., 2021). Additionally, the opaque, "black-box" nature of AI algorithms makes it difficult for both developers and users to understand how decisions are made (Rudin & Radin, 2019; Jo et al., 2023). This complexity, coupled with their ability to elicit strong trust responses, underscores the necessity for a careful and ethical deployment strategy. Understanding human trust in these systems is critical to ensure that AI-driven chatbots are beneficial while minimizing potential risks and harms.

This review aims to specifically address trust in AI-driven chatbots, focusing on the unique trust dynamics within this rapidly expanding field. By narrowing the scope, this study seeks to provide more generalizable findings and offer both practical and theoretical insights that are crucial for advancing AI chatbots. Such a focused review is essential to understand and mitigate the risks associated with these systems, particularly given their distinctive capabilities and the deep levels of user interaction they entail. Specifically, we ask:

RQ1: How is trust in AI-driven chatbots (a) conceptualised and (b) operationalised?

RQ2: What are the methodologies used to study trust in AI-driven chatbots?

RQ3: What factors influence trust development in AI-driven chatbots?

RQ4: What are the outcomes of trust in AI-driven chatbots?

3. Method

The inclusion and exclusion criteria, search term development, and search strategy are stated below building upon past systematic reviews in similar domains (Bach et al., 2022; Laranjo et al., 2018; Rheu et al., 2021; Zhang et al., 2023). The systematic review was conducted and reported in line with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 standards (Page et al., 2021). The search was conducted in 11 databases that are well-known digital libraries within the field of social science and human-computer interaction. The data was collected in August 2023.

Search terms: Two sets of keywords were searched: (1) Trust: "trust" OR "trustworthy" OR "trustworthiness"; (2) Machines: "intelligent agent" OR "conversational agent" OR "conversational system" OR "communicative agent" OR "interactive agent" OR "virtual assistant" OR "voice assistant" OR "dialogue system" OR "dialog system" OR "relational agent" OR "chatbot" OR "digital assistant".

Inclusion criteria: (1) focus on AI-driven chatbots that employ machine learning or natural language processing techniques to interpret user input and deliver unrestricted, unscripted, human-like, verbal or textual responses, distinct from their rule-based counterparts that rely on logical decision-making to offer predetermined responses, (2) the AI-driven chatbot is without physical embodiment, (3) the presence of AI is communicated explicitly, (4) an

empirical study that examines trust concepts in AI-driven chatbots, (5) user input was explicitly said to be unrestricted, (6) evaluate the AI application with regards to users (e.g., user's perception and evaluation of trust), (7) written in English, (8) peer-reviewed articles, conference proceedings, pre-prints, or dissertations, (9) the full-text is available, (10) published between January 2013 to July 2023.

Exclusion criteria: (1) studies using the Wizard of Oz method, (2) studies where users are told to imagine interactions with an AI-powered machine, such as studies that use vignettes as experimental stimuli, (3) studies of systems where user input occurred by clicking or tapping an answer amongst a set of predefined choices, (4) the article was an abstract/extended abstract, (5) the article was secondary research or focused on conceptual, theoretical, or methodological research.

4. Results

The initial search yielded 5259 results from 11 databases. 750 duplicates were removed using Rayyan.ai, an AI-powered systematic review tool to support title and abstract screening (Harrison et al., 2020). 4509 results remained for title and abstract screening. Two reviewers independently assessed 450 articles (10% of the total) based on titles and abstracts, achieving high inter-rater reliability with a Cohen's kappa coefficient of 0.758. The remaining 4059 articles were divided between the reviewers, who held regular meetings to validate inclusions/exclusions and resolve disagreements. 532 articles proceeded to full-text screening. Of these, 107 (20% of the remaining) were independently reviewed by the two experts. A high inter-rater reliability with a Cohen's kappa coefficient of 0.73 was achieved. They then divided the remaining 425 articles for full-text screening. Ultimately, 40 articles that met all inclusion criteria were included (Figure 1), and each article was assigned a number from A1 to A40 (Table 1 provides an overview).

General description

Scholarly interest in the topic has experienced significant growth through the years. No articles related to the research topic were found before 2017. One relevant article was identified in 2018, and this number increased to three in 2019 and another three in 2020. Notably, there was a substantial growth in research interest in 2021, with the identification of 15 articles. However, only four articles were found in 2022. In 2023, 14 articles were found, although it is important to note that the data collection for this year only extended until August, and more articles may have been published afterwards.

The geographical location where data collection occurred in the included articles had some diversity. Most of the included articles were based in the United States (n=11, 27.5%). Four articles (10%) were based in China, while three articles (7.5%) were based in the United Kingdom. Six articles (15%) collected data from multiple countries, while seven articles (17.5%) did not provide geographical information. The remaining nine articles collected data from nine different countries.

We also classified the articles' field of study based on the article content, publication venue, or first author's association. Most of the articles were conducted in the field of human-computer interaction (n=17, 42.5%). This field can be described by studies that examine how users interact with the machine, emphasising user experience and machine design. 12 articles (30%) were in the field of business or marketing, where the focus is on customer experience and engagement for business outcomes. Seven articles (17.5%) were in the field of

information systems, information technology, or information management. The articles in this field are focused on the managerial and organizational implications of the machine. Two articles (5%) were conducted in the field of psychology, focusing on psychological outcomes or psychological factors. The remaining two articles (5%) were in healthcare and encompass both physical and mental health.

There is some variation in the type of machine studied. The most common were voice or virtual assistants, like Siri, and smart speakers, like Google Home (n=29, 72.5%). We did not distinguish between voice assistants and smart speakers because the underlying natural language processing capabilities that power the machine's conversational ability are the same. Five articles (12.5%) studied text-based chatbots like Replika, three articles (7.5%) studied ChatGPT, two articles (5%) used chatbots developed by the researcher, while the last article (2.5%) broadly studied natural language generators like machine translators.

Only seven articles (17.5%) used study designs where participants had real-time interaction with the machine during the study. The other 33 articles (82.5%), that did not ensure participants' interaction with the machine during the study, used strategies like prompting participants to recall past conversations with the machines.

Most of the articles studied trust in the machine as an object (n=36, 90%). Only two articles (5%) studied trust in the manufacturer or producer of the machine, while the last two articles (5%) studied trust in the machine and trust towards the manufacturer.

RQ1a. Conceptualisation

Table 2 provides a list of the trust definitions found across the 40 included articles. Among these, 11 articles (27.5%) did not offer specific trust definitions, which were coded due to the lack of an explicit theoretical definition. The authors may have provided an implicit definition of trust in their measurements or operationalisations of trust, which will be further explored in RQ1b. Six articles (15%) introduced their own interpretations, while the remaining 23 (57.5%) referenced definitions from other researchers. Mayer et al. (1995) were cited in five articles, while Madsen and Gregor (2000) were cited in two. The fact that more than half of the articles adopted established definitions indicates a certain level of interdisciplinary consistency in how trust is defined. However, there is still significant diversity in the definitions used, as there is little repetition in the cited researchers. This suggests that trust remains a multifaceted concept within this research space. Despite the diversity in the definitions cited, common themes emerged in how trust was conceptualised across these articles. These themes were identified by coding each definition based on its dimension of trust, and then grouping similar dimensions into themes. Each definition can have multiple codes and hence, multiple dimensions. Figure 2 provides an illustration of the distribution of codes, dimensions, and themes. We term these themes as *trustor's behavioural state, trustor's emotional state, trustee's attributes,* and *situational circumstances*.

The theme of *trustor's behavioural state* describes definitions that emphasise the tangible and observable outcomes of trusting the AI-driven chatbot. This theme consists of 11 codes and four dimensions. Definitions in this theme include words like "rely," "depend," and "confide." Phrases that suggest a "willingness" to adopt an action, such as the willingness to act on the machine's recommendations, were also observed. The theme of *trustor's emotional state* refers to the user's affective involvement with the AI-driven chatbot. This theme consists of 16 codes and three dimensions. Definitions in this theme consider how users are "vulnerable," have "confidence" in the machine, and feel "safe" in their interactions

with the machine. *Trustee's attributes* refer to specific characteristics of the AI-driven chatbot that determine their trustworthiness. This theme consists of 20 codes and eight dimensions. The most common attributes are "benevolence," "integrity," "ability," or "competence"—concepts first introduced by Mayer et al. (1995), although not all these articles cited Mayer et al. (1995). Because all mentions of these attributes occurred together, they jointly constitute one dimension, rather than four dimensions. Whether machines are "reliable," "credible," "trustworthy," "sincere," or "safe" are also important attributes to describe trust in AI-driven chatbots. These machines are also evaluated based on characteristics of "efficiency" and "effectiveness." Lastly, definitions classified as *situational circumstances* refer to conditions that require an assessment by the user. This theme consists of six codes and three dimensions. It includes judgements of whether "privacy" and "security" are maintained, as well as whether "risk" is involved.

RQ1b. Operationalisation

Although there is no explicit theoretical definition, some articles may still provide an implicit definition of trust through their operationalisations. Thirty-five articles that employed the use of survey questionnaires in their study offered operationalisations of trust. Seven articles did not break down the concept of trust in their operationalisations (e.g., "I trust Alexa to make decisions" (A6)). This is concerning because trust is an abstract and multifaceted construct that different people might have differing views of. Hence, it becomes challenging to precisely understand the construct and its predictors without a detailed breakdown. One article used the single-dimensional construct of "dependable" to operationalise trust (A26). A total of 27 articles used multi-dimensional constructs to operationalise trust. Of the 27, 14 articles used multi-dimensional constructs but also used "trust" or "trustworthy" in their operationalisation. From the articles that used multi-dimensional constructs to operationalise trust, a total of 21 constructs were found and summarized in Figure 3.

The most common constructs were "ability/competence," "benevolence," and "integrity." This is aligned with the most common attributes used to define trust in terms of the trustee's attributes. "Reliable" was also a common construct used in the operationalisation of trust. Because multi-dimensional approaches offer richness and depth in capturing the multifaceted nature of trust, they are preferable to articles that do not break down the concept of trust.

We also checked the construct validity by assessing the extent to which the operationalisations reflect the theoretical definition of trust provided by the articles. Construct validity was only assessed for articles that defined and operationalised trust, but excluded articles that did not break down the concept of trust in their operationalisations (i.e., did not use multi-dimensional constructs to operationalise trust). From the 24 eligible articles, 13 articles used operationalisations that were completely misaligned from their chosen theoretical definition. This is concerning because the validity of any conclusions drawn about trust in the context of AI-driven chatbots may be compromised if the operationalisation does not accurately reflect the theoretical definition. Seven articles had some degree of construct validity, whereby their definition and operationalisation had some overlaps but were not perfect fits. Four articles (A5, A7, A12, A15) achieved construct validity, which meant that their chosen dimensions of trust were perfectly aligned with their theoretical definitions.

RQ2. Methodology

There is some diversity in the way researchers are studying trust (Table 2). The most common research method was cross-sectional surveys, with 32 articles (80%) employing surveys in their study. The second most common method was interviews with 11 articles (27.5%) conducting interviews with participants. Seven articles (17.5%) used cross-sectional surveys in combination with interviews. Other quantitative methods that incorporated the use of survey questionnaires with another method are experiment designs (A20, A26) and workshops (A3). The diversity of research methods used suggests that the field recognises the complex and multifaceted nature of trust. Researchers are adopting a range of tools and approaches to capture the nuances of trust in human-machine interactions. In addition to interviews, other diverse qualitative methods include crowdsourcing user experiences (A1), diary studies (A2, A37), user reviews on app stores (A21), and content analysis of news articles, blogs, and consumer reviews (A40). These are user-centric methods that indicate a focus on practical applications and the real-world impact of trust in AI-driven chatbots, which should be encouraged considering how entrenched these technologies are in our daily lives. Table 3 describes the sample size information for the studies that involved human participants.

Only A34 used a longitudinal survey—that lasted three weeks—in addition to a cross-sectional survey and interviews. Diary studies lasted five days (A37) and one week (A2). While these are preferable to methods that study trust at one single time point, the relatively short duration might not be sufficiently long to capture the evolution of trust in AI-driven chatbots.

RQ3. Process

A total of 27 articles investigated factors that contributed to trust development in AI-driven chatbots (Table 4). These factors were identified based on the hypotheses and variables tested by the included articles, or by inferring from the descriptions provided in qualitative articles. To provide a complete summary of the commonly studied factors in this research area, we included factors even if they had a statistically non-significant relationship with trust in the included articles. The factors were organized together in related categories, then related themes, based on their similarities. We further split RQ3 into two parts to better understand the conditions that influence trust development in AI-driven chatbots.

RQ3a. Predictors

These factors are *user-related factors, machine-related factors, interaction-related factors, social-related factors,* and *context-related factors.*

User-related factors refer to specific characteristics of the trustor or user. This theme can be further broken down into six categories of factors. *(1) Demographics*. Two articles investigated how age and gender contribute to trust in AI-driven chatbots. However, the results were non-significant. *(2) Knowledge*. Articles in this category investigated how different knowledge types affected trust in machines. Programming knowledge (A3) and the user's field of occupation (A30) had non-significant effects. However, A30 found a negative impact of information and communications technology expertise on trust. Having subjective knowledge in a topic area that the AI-driven chatbot is meant to help you with also hurts trust (A26). Prior experience had a positive impact (A15), but familiarity—which is a result of having prior experience—was non-significant (A38). *(3) Personality.* Articles in this category explored the relationship between personality dimensions, including the HEXACO personality dimensions (A7) and the big five personality traits (A15, A16, A25), and their

predictive role in trust. However, the findings were inconsistent, and a consensus has not been reached. Intellect had a non-significant impact on trust (A15, A16), and the same was observed for emotional instability (A15). Notably, state anxiety positively predicted trust (A16). *(4) Attitude.* A trusting stance and faith in general technology positively predicted trust (A25). Trust in the manufacturer (A13) and privacy cynicism, which represents a negative attitude towards the machine's handling of personal data (A35) had positive impacts on trust. Despite being unexpected, the authors argued that privacy cynicism serves as a coping mechanism, allowing consumers to overlook their privacy concerns. Learning style and immersive tendencies showed a non-significant impact on trust (A16), and similar non-significant results were observed for technology optimism (A38), affinity for technology interaction (A33), and personal innovativeness (A28). *(5) Perception.* Perceptions of machine human-likeness, including voice humanness (e.g., natural pronunciation) and understanding humanness (e.g., avoiding misunderstandings), produced mixed effects on trust (A12). The social categorisation of the machine concerning its gender, age, race, and professional identity also yielded mixed effects on trust (A37). Notably, the perception of an authentic experience was a positive predictor of trust (A19). *(6) Cognitive.* The evaluation of risks that make users feel vulnerable when adopting the machine negatively predicted trust (A6). Privacy risks, in particular, negatively influenced trust (A29, A33). However, privacy concerns yielded inconsistent results, with some studies reporting non-significant associations (A5, A28) and others indicating a positive prediction of trust (A15).

Machine-related factors refer to specific characteristics of the trustee or machine, they are directly related to the machine's capabilities. This theme can be further broken down into three categories of factors. *(1) Affective.* This refers to the emotional aspects of the machine. A machine that was seen as sociable or friendly (A28), caring and non-judgemental (A27), or had emotional competency to recognise and respond appropriately to the user's emotions (A36) all positively predicted trust. However, a machine's relational competency in terms of maintaining a harmonious relationship with the user had a non-significant impact on trust. *(2) Behavioural.* This refers to the observable behaviours of the machine. An anthropomorphic machine that is ascribed with human-like characteristics positively predicted trust (A13, A18, A33). The same positive effect is observed for speaking humanness and listening humanness (A34). Behavioural traits of the machine like being up-to-date and well-mannered also positively predicted trust (A33). *(3) Practical.* This refers to the pragmatic and functional aspects related to the machine. A competent machine (A5, A26, A36) or a machine with high information quality (A32) or high system quality (A32, A38) positively predicted trust. A machine that ensures data security (A27, A39) and one that allows for personalised interactions (A19, A24) exhibited a positive predictive effect on trust. From this, it is evident that trust is deeply reliant on the human-like characteristics of the machine both in terms of the emotional and observable dimensions, and the high-functioning abilities of the machine.

Interaction-related factors are derived from user interactions with the machine. These factors can only be observed by involving both the user and the machine. Machine failures, indicated by incorrect actions or failure to respond, negatively predicted trust (A1). Conversely, perceived ease of use positively predicted trust (A5, A28). Factors such as perceived auditory control over the machine (A20), feelings of relationship closeness with the machine (A13), task efficiency (A24), and social presence (A5) all positively predicted trust. However, usage barriers (A6), information seeking (A24), and playfulness (A24) did not yield a statistically significant effect on trust. Notably, the perceived usefulness and perceived enjoyment of the machine had conflicting results. While A5 found non-significant effects for

both, A28 reported a positive predictive effect of perceived usefulness on trust, and A38 found that perceived enjoyment positively predicted trust.

Social-related factors refer to factors that involve other people beyond the direct interactions between the user and the machine. Electronic word-of-mouth, arising from social interactions where individuals share their opinions about the machine, had a positive predictive effect on trust (A28). Similarly, subjective norms, reflecting the social pressure a person feels regarding machine adoption, positively predicted trust (A38). The perception of the machine as a means to gain symbolic rewards, known as status seeking, also exhibited a positive predictive effect on trust (A6). However, the degree of social interaction, specifically the extent to which individuals use the machine to communicate and interact with others, did not yield a statistically significant effect on trust (A24).

Context-related factors refer to conditions surrounding the usage of the machine. Factors like the convenience offered by the machine to enrich the online shopping experience (A6) and whether the user purchased the machine for themselves (A13) positively predicted trust. However, factors such as ownership duration, number of machines owned, and whether the user used their personal account had a non-significant effect on trust (A13).

RQ3b. Mediators and moderators

Only three articles studied the mediators of trust, and none studied the moderators. Mediators elucidate the mechanisms through which trust operates, offering insights into the internal processes that connect different factors to trust. On the other hand, moderators identify the conditions that influence the strength or direction of the relationship between different factors and trust. The dearth of studies exploring the processes and boundary conditions of trust is concerning because it leaves a significant gap in our understanding of the contextual factors that shape trust in AI-driven chatbots.

All the identified mediators were categorised as interaction-related factors. Task attraction, which refers to the perceived ability of the machine to complete given tasks, and social attraction, which refers to users' willingness to engage in friendly communication with the machine, positively mediated the relationship between anthropomorphism and trust across both cognitive and emotional dimensions (A18). Additionally, interaction quality was identified as a positive mediator between information and system quality and trust (A32). This underscores the importance of considering both interaction-related and machine-related factors in the development of trust. Furthermore, social presence, encompassing the warmth, sociability, and sense of intimacy conveyed by the machine, was revealed as a mediator in the congruency effect of speaking humanness and listening humanness on trust (A34).

RQ4. Outcomes

A total of 27 articles investigated the outcomes of trust in AI-driven chatbots. These factors were identified in a similar manner as the factors in RQ3 (Table 5). We included both direct and indirect outcomes of trusting AI-driven chatbots, as well as outcomes where trust played a moderating effect. Sixteen articles studied the direct outcomes of trust, nine articles studied both direct and indirect outcomes, and two articles studied the moderating role of trust on an outcome. Indirect outcomes suggest a potential mediator between trust and an outcome, though a formal mediation analysis was not conducted in all but one of the articles. Examples of these potential mediators include satisfaction with the machine and engagement

with the machine. The outcomes were thematically organized into *affective outcome, relational outcome, behavioural outcome, cognitive outcome,* and *psychological outcome.*

Affective outcomes refer to changes in feelings and attitudes towards the machine. All four articles that examined shifts in attitude toward the machine reported a positive correlation with trust (A5, A9, A28, A33). The three articles that examined satisfaction with the machine also reported a positive correlation with trust (A4, A11, A36). Given that emotional experiences significantly impact other types of judgments, none of the articles isolated affective outcomes; instead, they examined the interconnected nature of affective responses with cognitive, behavioural, and relational aspects.

Relational outcomes refer to changes in the relationship with the machine or with the manufacturer of the machine. In terms of brand loyalty for the manufacturer, A10 found a positive effect on user trust. A similar positive effect was found for relationship quality with the machine (A17) and on customer experience (A11, A15). These positive effects underscore the significance of trust in influencing customer sentiments, which has implications that are particularly important for studies in the business or marketing disciplines that are working towards customer-oriented goals.

Behavioural outcomes refer to the perceptible changes in how a user acts because of trusting the machine and were the most common form of outcome investigated. Among these, behaviours associated with the use of the machine were the most frequently examined by 18 articles, with trust demonstrating a positive influence on all outcomes related to usage. These outcomes include willingness to use the machine, intention to use the machine, continuance usage intentions, habit, and actual adoption. Four articles observed a positive effect of trust on the self-disclosure of personal information to machines (A2, A9, A27, A31). Three articles observed a positive effect of trust on engagement with the machine (A9, A17, A36). A4 found that trust in the machine had a positive effect on engagement with work and productivity at work. A11 observed that trust had a positive impact on user recommendations, whereby users are more likely to share their positive experiences with others in their social circle if they trust the machine. A40 found that trust moderates the usage of the machine for task function and information search, which encompasses activities like using the machine to control Internet of Things devices and searching for product information. These findings collectively underscore the broad impact of trust on diverse user behaviours and interactions with AI-driven chatbots.

Cognitive outcomes relate to thinking and reasoning and refer to a change in the mental processes used in evaluating the machine. Trust moderated the relationship between interactivity and perceived performance (A22), which suggests that a higher level of trust strengthens the relationship between the interactivity dimensions and the capabilities of the machine. Higher levels of trust in the technology positively impacted performance expectancy, indicating that users with greater trust expected superior performance from the technology (A29). Additionally, trust had positive effects on involvement and self-brand connection, signifying that the machine held more significance for users, and the brand became more integrated into consumers' self-concepts (A19). However, trust hurts perceived risks, which is the extent to which people believe that there will be a potential loss in the event of a release of personal information (A31). This emphasises the delicate balance between user trust and evaluations of the machine.

Psychological outcomes concern the emotional state of users. According to A8, trust in the machine can benefit self-esteem and psychological well-being, which may stem from

the sense of control and empowerment individuals experience when they use the machine. However, A17's findings revealed that trust can lead to psychological dependence. This is demonstrated by manifestations like salience, tolerance, and withdrawal—well-documented signs associated with addiction; a phenomenon often observed in online gaming. This draws attention to the importance of understanding post-adoption behaviours and the possibility of negative repercussions to trusting AI-driven chatbots.

5. Discussion

This systematic review synthesized the conceptualisations and operationalisations of trust in AI-driven chatbots, the methodologies used to study it, the process of trust development, and outcomes of trust in AI-driven chatbots. Below, we discuss the findings and implications of the review and summarise future research suggestions in Table 6.

Ensuring clarity and consistency in trust conceptualisation and operationalisation

Four overarching themes of definitions emerged, revealing the diverse ways researchers conceptualise trust in AI-driven chatbots. These conceptualisations were drawn from user behaviour and emotions, machine attributes, and situational contexts. However, it was also observed that a significant number of articles failed to provide definitions of trust. This echoes the findings of other reviews of AI systems beyond AI-driven chatbots (Bach et al., 2022; Ueno et al., 2022), suggesting that the absence of clear definitions is prevalent across the broader field. The absence of clear and explicit definitions poses a challenge because a precise understanding of trust is fundamental for both research and practical applications. Without a well-defined conceptualisation of trust, researchers and practitioners may encounter difficulties in interpreting and comparing findings across studies, potentially limiting the advancement of knowledge in this field.

To establish an appropriate definition of trust, researchers need to align their choice with the specific objectives of their study. For example, investigations aiming to unravel the impact of user characteristics on trusting beliefs will benefit from a definition centred around user behaviour or emotional states. On the other hand, studies exploring the influence of machine attributes on trust should opt for a definition that centres on those specific attributes. The choice of one conceptualisation over another holds significant implications for how trust is assessed, subsequently influencing the study's outcomes (Zhang et al., 2023). This approach can be likened to Yang and Wibowo's (2022) conclusion to pick evaluation criteria that are suitable for the specific AI application or Bach et al.'s (2022) suggestion to pick a trust definition that is most appropriate for the context.

Furthermore, several articles failed to provide clear operationalisations of trust. Studies that do not dissect trust have a heightened risk of conducting an ambiguous assessment, potentially compromising the study's ability to achieve its intended outcomes. It follows that the operationalisations of trust within these studies should carefully assess and capture the pertinent dimensions aligned with their defined conceptualisation. The lack of construct validity in existing studies, which can also be observed in studies on trust in social media (Zhang et al., 2023), warrants heightened attention. A misalignment in definitions and operationalisations can compromise the validity and reliability of study findings, potentially leading to inaccurate or incomplete conclusions about trust in these systems. Hence, future studies should ensure methodological rigour by aligning definitions with operationalisations.

Examining trust development via longitudinal studies

Moreover, our review revealed that cross-sectional surveys remain the dominant methodology, with only three articles employing a longitudinal study design. Given that trust development is a complex and dynamic process, it is crucial to examine its evolution over time, capturing key stages such as its initial establishment, maintenance, enhancement, collapse, and reconstruction (Langer et al., 2023; Skjuve et al., 2021). This temporal dimension of trust allows researchers to uncover the factors that influence trust at each stage, providing a more comprehensive understanding of trust dynamics in human-machine interactions. Yang and Wibowo (2022) similarly emphasized the need for longitudinal studies, noting their importance in understanding how trust in AI affects long-term outcomes, such as customer lifetime value. Moreover, longitudinal studies enable the exploration of critical challenges that arise during prolonged interaction, such as how machine performance, failures, or ease of use affect trust. This understanding is vital for identifying factors that predict trust erosion or recovery, thereby informing the design of trustworthy AI systems.

Longitudinal studies may also address the "novelty effect," a phenomenon where users initially exhibit high levels of trust due to the excitement and curiosity of interacting with new technology (Miguel-Alonso et al., 2024). This effect often skews assessments of trust during the early stage, but trust may decline as the excitement fades or if performance issues arise. This phenomenon has been observed in various forms of human-machine interaction (Heyselaar, 2023; Miguel-Alonso et al., 2024). The temporal aspect becomes crucial here as it allows researchers to distinguish between the impact of the novelty effect and the genuine development of trust over time. Longitudinal studies would provide a valuable opportunity to observe how trust levels fluctuate beyond the initial encounter.

Elucidating the processes and boundary conditions of trust

The findings indicate that existing work examined a myriad of antecedents of trust in AI-driven chatbots, categorisable into five themes, including user, machine, interaction, social, and context-related factors. A significant portion of the articles focused on predictors of trust, rather than mediators or moderators. Similarly, existing reviews fail to distinguish between the predictors, mediators, and moderators of trust (Rheu et al., 2021; Yang & Wibowo, 2022). While identifying predictors is essential, understanding the intricate dynamics of trust development requires exploring not only what influences trust but also the intermediary processes (mediators) and boundary conditions (moderators) that shape this complex psychological phenomenon. Only three articles were identified to have examined the mediators of trust. Therefore, there is a need for more comprehensive research that considers a broader spectrum of mediators to enrich our understanding of the nuanced mechanisms behind trust in this context. Future research in this domain could benefit from addressing this notable gap by adopting a more nuanced approach that considers the multifaceted aspects influencing the development of trust in AI-driven chatbots.

Addressing cultural and demographic diversity in trust development

Additionally, a critical aspect often overlooked in studies of trust development in AI-driven chatbots is the role of cultural and demographic diversity. Current studies predominantly stem from specific regions, particularly Western countries such as the United States. This raises concerns regarding the generalizability of findings, as people from different cultures may possess varying conceptions and expectations of trust in human-chatbot interactions. For example, trust dynamics may be shaped by culturally specific

norms, values, and communication styles (Yuki et al., 2005) that influence how users perceive and interact with AI-driven systems (Liu et al., 2024). Similarly, demographic diversity—including age, educational background, and digital literacy—may serve as boundary conditions that influence the trajectory of trust development (Bentley et al., 2024). Future research should expand the diversity of the sample population, especially to include users from different socioeconomic backgrounds and levels of digital literacy, to obtain more representative data and conclusions. Conducting cross-cultural comparative studies and expanding the diversity of sample populations will provide a more comprehensive picture of trust dynamics.

Understanding benefits and risks of trust in AI chatbots

By looking at both direct and indirect outcomes of trust, the results reveal that the outcomes of trust in AI-driven chatbots can be organized into five overarching themes, including affective, relational, behavioural, and cognitive outcomes. This provides valuable insights into the implications of users' interactions with AI-driven chatbots. The findings underscore the existence of a relatively understudied outcome: psychological outcomes concerning the emotional state of users. Psychological outcomes are overlooked even in the broader field of AI studies. Yang and Wibowo's (2022) review which focuses on broader AI applications uncovered outcomes of trust in AI that were categorised as cognitive, affective, and behavioural changes—similarly neglecting psychological outcomes. Only two articles in the corpus studied three types of psychological outcomes, examining the impact of trust on self-esteem, psychological well-being, and dependence. The inclusion of dependence as a psychological outcome is of particular significance, given its association with internet addiction and pathological internet use (Widyanto & Griffiths, 2006). It is crucial to recognise that excessive internet use is often rooted in underlying psychopathological conditions and maladaptive cognitions (Widyanto & Griffiths, 2006). The implications become especially pertinent when considering the application of AI-driven chatbots in mental healthcare (Miner et al., 2016; Vaidyam et al., 2019). The convergence of AI-driven chatbots with mental health raises concerns about the potential unintended consequences when employing this technology to address mental health concerns, particularly for a potentially vulnerable group of users. Thus, there is a need for in-depth exploration of these psychological ramifications. Understanding these benefits and risks of trust in AI chatbots is crucial for harnessing their potential while mitigating adverse effects, ensuring that they contribute positively to user well-being and do not exacerbate existing vulnerabilities.

6. Limitations

One limitation is related to the scope of investigation into the trust dynamics. Specifically, only a limited number of articles delved into the examination of trust in the manufacturer of the machine, while the majority focused on the machine alone. This limitation underscores a gap in the understanding of how users differentiate their trust between the AI system and the entity responsible for its creation. This dual perspective is crucial for comprehensively grasping the intricacies of trust in AI-driven chatbots, as users' trust might be influenced not only by the machine's performance but also by their perceptions of the manufacturer's reliability and ethical considerations.

Another limitation of this study is the reliance on recall-based studies for the review, with only 17.5% of the included articles involving real-time interactions between participants and AI-driven chatbots. While recall-based studies reflect users' actual experiences retrospectively, they are subject to potential biases in participants' memories and the

influence of retrospective interpretation. Although vignette-based studies were excluded due to their reliance on hypothetical scenarios that may not fully capture the complexities of real interactions, their inclusion could provide additional insights by controlling for specific scenarios and offering a different perspective on users' anticipated interactions with AI. This consideration highlights the potential value of including vignette-based studies in future reviews to enhance the comprehensiveness of the findings.

Third, this study focused on the processes of building and maintaining trust in AI-driven chatbots, while providing limited attention to the concept of distrust. Although trust and distrust are closely related, they are distinct phenomena with different sets of expectations and manifestations (Saunders et al., 2014). Distrust can influence user behaviour with AI-driven chatbots just as significantly as trust. Future research should incorporate an examination of distrust to offer a more comprehensive understanding of the complexities surrounding trust in AI chatbots.

Last, while this study acknowledges the diversity in AI-driven chatbots, it does not address the potential impact of different chatbot types on user trust. Different chatbot formats may foster trust in distinct ways. For example, trust in voice assistants may be influenced by the chatbot's speaking and listening humanness (Hu et al., 2023), but trust in text-based chatbots cannot develop in the same manner due to the absence of these features. Future research could benefit from meta-analytic studies that quantitatively aggregate findings from various studies to provide a precise understanding of how different chatbot types affect trust.

7. Conclusion

This study aimed to illuminate how the concept of trust is studied in the HMC field. Through an exploration of 40 articles, this systematic review outlined methodological trends, identified patterns in the conceptualisation and operationalisation of trust, highlighted key factors influencing trust development, and explored the outcomes associated with trusting AI-driven chatbots. We provided recommendations aimed at addressing the research gaps in this space, which will be helpful given the ongoing evolution of the technology and the increasing scholarly interest in this burgeoning field. By shedding light on trust dynamics, this review provides a foundation for future research efforts aimed at mitigating potential risks and maximizing the benefits of AI technologies in human-machine interactions.

Bibliography

Acikgoz, F., & Vega, R. P. (2022). The Role of Privacy Cynicism in Consumer Habits with Voice Assistants: A Technology Acceptance Model Perspective. International Journal of Human-Computer Interaction, 38(12), 1138–1152. https://doi.org/10.1080/10447318.2021.1987677

Acikgoz, F., Perez-Vega, R., Okumus, F., & Stylos, N. (2023). Consumer engagement with AI-powered voice assistants: A behavioral reasoning perspective. Psychology and Marketing. https://doi.org/10.1002/mar.21873

Alimamy, S., & Kuhail, M. A. (2023). I will be with you Alexa! The impact of intelligent virtual assistant's authenticity and personalization on user reusage intentions. Computers in Human Behavior, 143. https://doi.org/10.1016/j.chb.2023.107711

Asan, O., Bayrak, A. E., & Choudhury, A. (2020). Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. Journal of medical Internet research, 22(6), e15154. https://doi-org/10.2196/15154

Ashrafi, D. M., & Easmin, R. (2022). Okay Google, Good to Talk to You... Examining the Determinants Affecting Users' Behavioral Intention for Adopting Voice Assistants: Does Technology Self-Efficacy Matter? International Journal of Innovation and Technology Management, 20(2). https://doi.org/10.1142/S0219877023500049

Bach, T. A., Khan, A., Hallock, H., Beltrão, G., & Sousa, S. (2022). A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective. International Journal of Human–Computer Interaction, https://doi.org/10.1080/10447318.2022.2138826

Baek, T. H., & Kim, M. (2023). Is ChatGPT scary good? How user motivations affect creepiness and trust in generative artificial intelligence. Telematics and Informatics, 83. https://doi.org/10.1016/j.tele.2023.102030

Baughan, A., Wang, X., Liu, A., Mercurio, A., Chen, J., & Ma, X. (2023). A Mixed-Methods Approach to Understanding User Trust after Voice Assistant Failures. Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/3544548.3581152

Bawack, R. E., Wamba, S. F., & Carillo, K. D. A. (2021). Exploring the role of personality, trust, and privacy in customer experience performance during voice shopping: Evidence from SEM and fuzzy set qualitative comparative analysis. International Journal of Information Management, 58. https://doi.org/10.1016/j.ijinfomgt.2021.102309

Bentley, S. V., Naughtin, C. K., McGrath, M. J., Irons, J. L., & Cooper, P. S. (2024). The digital divide in action: how experiences of digital technology shape future relationships with artificial intelligence. AI and Ethics, 4(4), 901–915. https://doi.org/10.1007/s43681-024-00452-3

Brzowski, M., & Nathan-Roberts, D. (2019). Trust Measurement in Human–Automation Interaction: A Systematic Review. Proceedings of the Human Factors and

Ergonomics Society Annual Meeting, 63(1), 1595–1599. https://doi.org/10.1177/1071181319631462

Cabrero-Daniel, B., & Cabrero, A. S. (2023). Perceived Trustworthiness of Natural Language Generators. Proceedings of the First International Symposium on Trustworthy Autonomous Systems. https://doi.org/10.1145/3597512.3599715

Chandra, S., Shirish, A., & Srivastava, S. C. (2022). To Be or Not to Be …Human? Theorizing the Role of Human-Like Competencies in Conversational Artificial Intelligence Agents. Journal of Management Information Systems, 39(4), 969–1005. https://doi.org/10.1080/07421222.2022.2127441

Chaturvedi, R., Verma, S., Das, R., & Dwivedi, Y. K. (2023). Social companionship with artificial intelligence: Recent trends and future avenues. Technological Forecasting and Social Change, 193, 122634–122634. https://doi.org/10.1016/j.techfore.2023.122634

Chen, Q. Q., & Park, H. J. (2021). How anthropomorphism affects trust in intelligent personal assistants. Industrial Management and Data Systems, 121(12), 2722–2737. https://doi.org/10.1108/IMDS-12-2020-0761

Choudhury, A., & Shamszare, H. (2023). Investigating the Impact of User Trust on the Adoption and Use of ChatGPT: Survey Analysis. Journal of Medical Internet Research, 25. https://doi.org/10.2196/47184

Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., Murad, C., Munteanu, C., Wade, V., & Cowan, B. R. (2019). What makes a good conversation? Challenges in designing truly conversational agents. Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/3290605.3300705

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). Online trust: Concepts, evolving themes, a model. International Journal of Human-Computer Studies, 58(6), 737–758.

Denecke, L., Abd-Alrazaq, A., & Househ, M. (2021). Artificial Intelligence for Chatbots in Mental Health: Opportunities and Challenges. In M. Househ, E. Borycki & A. Kushniruk (Eds.), Multiple Perspectives on Artificial Intelligence in Healthcare. Springer Cham.

Dietz, G., Den Hartog, D. N., Sanders, K., & Schyns, B. (2006). Measuring trust inside organisations. Personnel Review, 35(5), 557–588. https://doi.org/10.1108/00483480610682299

Ejdys, J. (2018). Building technology trust in ICT application at a University. International Journal of Emerging Markets, 13(5), 980–997.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. MIS Quarterly: Management Information Systems, 27(1), 51–90. https://doi.org/10.2307/30036519

Giddens, A. (1990). The Consequences of Modernity. Cambridge: Polity Press.

Guzman, A. L. (2020). Ontological boundaries between humans and computers and the implications for Human-Machine Communication. Human-Machine Communication, 1, 37-54. https://doi.org/10.30658/hmc.1.3

Harrington, C. N., & Egede, L. (2023). Trust, Comfort and Relatability: Understanding Black Older Adults' Perceptions of Chatbot Design for Health Information Seeking. Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/3544548.3580719

Harrison, H., Griffin, S. J., Kuhn, I., & Usher-Smith, J. A. (2020). Software tools to support title and abstract screening for systematic reviews in healthcare: an evaluation. BMC Medical Research Methodology, 20(1). https://doi.org/10.1186/s12874-020-0897-3

Hasan, R., Shams, R., & Rahman, M. (2021). Consumer trust and perceived risk for voice-controlled artificial intelligence: The case of Siri. Journal of Business Research, 131, 591–597. https://doi.org/10.1016/j.jbusres.2020.12.012

Heyselaar, E. (2023). The CASA theory no longer applies to desktop computers. Scientific Reports, 13(1). https://doi.org/10.1038/s41598-023-46527-9

Hoff, K. A., & Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. Human Factors, 57(3), 407–434. https://doi.org/10.1177/0018720814547570

Hsu, W.-C., & Lee, M.-H. (2023). Semantic Technology and Anthropomorphism: Exploring the Impacts of Voice Assistant Personality on User Trust, Perceived Risk, and Attitude. Journal of Global Information Management, 31(1). https://doi.org/10.4018/JGIM.318661

Hu, P., Gong, Y., Lu, Y., & Ding, A. W. (2023). Speaking vs. Listening? Balance conversation attributes of voice assistants for better voice marketing. International Journal of Research in Marketing, 40(1), 109–127. https://doi.org/10.1016/j.ijresmar.2022.04.006

Hu, P., Lu, Y., & Gong, Y. (2021). Dual humanness and trust in conversational AI: A person-centered approach. Computers in Human Behavior, 119. https://doi.org/10.1016/j.chb.2021.106727

Hu, X., Wu, G., Wu, Y., & Zhang, H. (2010). The effects of Web assurance seals on consumers' initial trust in an online vendor: A functional perspective. Decision Support Systems, 48(2), 407–418.

Jenneboer, L., Herrando, C., & Constantinides, E. (2022). The Impact of Chatbots on Customer Loyalty: A Systematic Literature Review. Journal of Theoretical and Applied Electronic Commerce Research, 17(1), 212–229. https://doi.org/10.3390/jtaer17010011

Jo, E., Epstein, D. A., Jung, H., & Kim, Y.-H. (2023). Understanding the Benefits and Challenges of Deploying Conversational AI Leveraging Large Language Models for Public Health Intervention. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3544548.3581503

Komiak, S. Y., & Benbasat, I. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. MIS Quarterly, 30(4), 941-960.

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2022). Interacting with educational chatbots: A systematic review. Education and Information Technologies, 28(1), 973–1018. https://doi.org/10.1007/s10639-022-11177-3

Kumar, N. (1996). The power of trust in manufacturer-retailer relationships. Harvard Business Review, 74(6), 92. https://ink.library.smu.edu.sg/lkcsb_research/5179/

Langer, M., König, C. J., Back, C., & Hemsing, V. (2023). Trust in Artificial Intelligence: Comparing Trust Processes Between Human and Automated Trustees in Light of Unfair Bias. Journal of Business and Psychology, 38(3), 493–508. https://doi.org/10.1007/s10869-022-09829-9

Lankton, N. K., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. Journal of the Association for Information Systems, 16(10), 880–918.

Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S., & Coiera, E. (2018). Conversational agents in healthcare: a systematic review. Journal of the American Medical Informatics Association, 25(9), 1248–1258. https://doi.org/10.1093/jamia/ocy072

Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. Proceedings of the ACM on Human-Computer Interaction, 2. https://doi.org/10.1145/3274371

Lee, O.-K. D., Ayyagari, R., Nasirian, F., & Ahmadian, M. (2021). Role of interaction quality and trust in use of AI-based voice-assistant systems. Journal of Systems and Information Technology, 23(2), 154–170. https://doi.org/10.1108/JSIT-07-2020-0132

Lee, S. K., & Sun, J. (2022). Testing a theoretical model of trust in human-machine communication: emotional experience and social presence. Behaviour & Information Technology, 1–14. https://doi.org/10.1080/0144929x.2022.2145998

Lee, T. (2005). The impact of perceptions of interactivity on customer trust and transaction intentions in mobile commerce. Journal of Electronic Commerce Research, 6(3), 165.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Li, X., Hess, T. J., & Valacich, J. S. (2008). Why do we trust new technology? A study of initial trust formation with organizational information systems. The Journal of Strategic Information Systems, 17(1), 39-71.

Li, H., & Zhang, R. (2024). Finding Love in Algorithms: Deciphering the Emotional Contexts of Close Encounters with AI Chatbots. Journal of Computer-Mediated Communication.

Liu, Y., Gan, Y., Song, Y., & Liu, J. (2021). What Influences the Perceived Trust of a Voice-Enabled Smart Home System: An Empirical Study. Sensors (Basel, Switzerland), 21(6), 2037. https://doi.org/10.3390/s21062037

Liu, Z., Li, H., Chen, A., Zhang, R., Lee, Y.-C., Sas, C., Dugas, P. T., Wilson, M. L., Williamson, J. R., Shklovski, I., Kyburz, P., & Mueller, F. F. (2024). Understanding Public Perceptions of AI Conversational Agents: A Cross-Cultural Analysis. Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, 1–17. https://doi.org/10.1145/3613904.3642840

Luo, X., Li, H., Zhang, J., & Shim, J. P. (2010). Examining multi-dimensional trust and multi-faceted risk in initial acceptance of emerging technologies: An empirical study of mobile banking services. Decision Support Systems, 49(2), 222–234.

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human-human and human-automation trust: An integrative review. Theoretical Issues in Ergonomics Science, 8(4), 277–301. https://doi.org/10.1080/14639220500337708

Madsen, M., & Gregor, S.D. (2000). Measuring Human-Computer Trust. In Proceedings of the 11th Australian Conference on Information Systems, 53-64.

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet Users' Information Privacy Concerns (IUIPC): The construct, the scale, and a causal model. Information Systems Research, 15(4), 336–355.

Malodia, S., Ferraris, A., Sakashita, M., Dhir, A., & Gavurova, B. (2023). Can Alexa serve customers better? AI-driven voice assistant service interactions. Journal of Services Marketing, 37(1), 25–39. https://doi.org/10.1108/JSM-12-2021-0488

Malodia, S., Islam, N., Kaur, P., & Dhir, A. (2021). Why Do People Use Artificial Intelligence (AI)-Enabled Voice Assistants? IEEE Transactions on Engineering Management. https://doi.org/10.1109/TEM.2021.3117884

Marikyan, D., Papagiannidis, S., Rana, O. F., Ranjan, R., & Morgan, G. (2022). "Alexa, let's talk about my productivity": The impact of digital assistants on work productivity. Journal of Business Research, 142, 572–584. https://doi.org/10.1016/j.jbusres.2022.01.015

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. The Academy of Management Review, 20(3), 709–734. https://doi.org/10.2307/258792

McAllister, D. J., Lewicki, R. J., & Chaturvedi, S. (2006). Trust in developing relationships: From theory to measurement. Academy of Management Proceedings, 2006(1). https://doi.org/10.5465/ambpp.2006.22897235

McKnight, D. H., & Chervany, N. L. (2001). What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology. International Journal of Electronic Commerce, 6(2), 35–60.

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. ACM Transactions on Management Information Systems, 2(2), 1-25.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. Information Systems Research, 13(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

Meyerson, D., Weick, K. E., & Kramer, R. M. (1996). Swift trust and temporary groups. In R. M. Kramer & T. R. Tyler (Eds.), Trust in organizations: Frontiers of theory and research (pp. 166–195). Sage Publications, Inc. https://doi.org/10.4135/9781452243610.n9

Miguel-Alonso, I., Checa, D., Henar Guillen-Sanz, & Bustillo, A. (2024). Evaluation of the novelty effect in immersive Virtual Reality learning experiences. Virtual Reality, 28(1). https://doi.org/10.1007/s10055-023-00926-5

Miner, A. S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., & Linos, E. (2016). Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health. JAMA Internal Medicine, 176(5), 619. https://doi.org/10.1001/jamainternmed.2016.0400

Müller, L., Mattke, J., Maier, C., Weitzel, T., & Graser, H. (2019). Chatbot acceptance: A latent profile analysis on individuals' trust in conversational agents. SIGMIS-CPR 2019 - Proceedings of the 2019 Computers and People Research Conference, 35–42. https://doi.org/10.1145/3322385.3322392

Oh, Y. J., Zhang, J., Fang, M.-L., & Fukuoka, Y. (2021). A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss. International Journal of Behavioural Nutrition and Physical Activity, 18(1). https://doi.org/10.1186/s12966-021-01224-6

Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. Computers and Education. Artificial Intelligence, 2, 100033–100033. https://doi.org/10.1016/j.caeai.2021.100033

Oliveira, G. G. D., Lizarelli, F. L., Teixeira, J. G., & Mendes, G. H. D. S. (2023). Curb your enthusiasm: Examining the customer experience with Alexa and its marketing outcomes. Journal of Retailing and Consumer Services, 71. https://doi.org/10.1016/j.jretconser.2022.103220

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & McGuinness, L. A. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Systematic Reviews, 10(1). https://doi.org/10.1186/s13643-021-01626-4

Pal, D., Arpnikanondt, C., & Razzaque, M. A. (2020). Personal Information Disclosure via Voice Assistants: The Personalization–Privacy Paradox. SN Computer Science, 1(5). https://doi.org/10.1007/s42979-020-00287-9

Pal, D., Babakerkhell, M. D., & Zhang, X. (2021). Exploring the Determinants of Users' Continuance Usage Intention of Smart Voice Assistants. IEEE Access, 9, 162259–162275. https://doi.org/10.1109/ACCESS.2021.3132399

Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. International Journal of Electronic Commerce, 7(3), 101–134.

Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of replika. Computers in Human Behavior, 140, 107600–107600. https://doi.org/10.1016/j.chb.2022.107600

Pfeuffer, N., Adam, M., Toutaoui, J., Hinz, O., & Benlian, A. (2019). Mr. And MRS. Conversational agent—Gender stereotyping in judge-advisor systems and the role of egocentric bias. 40th International Conference on Information Systems, ICIS 2019. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85114902431&partnerID=40&md5=85c263afdc61c37dbbdfcd0ab1956fb9

Pitardi, V., & Marriott, H. R. (2021). Alexa, she's not human but… Unveiling the drivers of consumers' trust in voice-based artificial intelligence. Psychology and Marketing, 38(4), 626–642. https://doi.org/10.1002/mar.21457

Poushneh, A. (2021). Impact of auditory sense on trust and brand affect through auditory social interaction and control. Journal of Retailing and Consumer Services, 58. https://doi.org/10.1016/j.jretconser.2020.102281

Prahl, A., & Van Swol, L. M. (2021). Out with the humans, in with the machines?: Investigating the behavioral and psychological effects of replacing human advisors with a machine. Human-Machine Communication, 2, 209–234. https://doi.org/10.30658/hmc.2.11

Prakash, A., & Das, S. (2020). Intelligent Conversational Agents in Mental Healthcare Services: A Thematic Analysis of User Perceptions. Pacific Asia Journal of the Association for Information Systems, 12(2), 1–34. https://doi.org/10.17705/1pais.12201

Purwanto, P., Kuswandi, K., & Fatmah, F. (2020). Interactive Applications with Artificial Intelligence: The Role of Trust among Digital Assistant Users. Foresight and STI Governance, 14(2), 64–75. https://doi.org/10.17323/2500-2597.2020.2.64.75

Raiche, A.-P., Dauphinais, L., Duval, M., De Luca, G., Rivest-Hénault, D., Vaughan, T., Proulx, C., & Guay, J.-P. (2023). Factors influencing acceptance and trust of chatbots in juvenile offenders' risk assessment training. Frontiers in Psychology, 14. https://doi.org/10.3389/fpsyg.2023.1184016

Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. International Journal of Human-Computer Studies, 151, 102630–102630. https://doi.org/10.1016/j.ijhcs.2021.102630

Rheu, M., Shin, J. Y., Peng, W., & Huh-Yoo, J. (2021). Systematic Review: Trust-Building Factors and Implications for Conversational Agent Design. International Journal of

Human–Computer Interaction, 37(1), 81-96. https://doi-org/10.1080/10447318.2020.1807710

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. Journal of Personality, 35(4), 651–665. https://doi.org/10.1111/j.1467-6494.1967.tb01454.x

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. Academy of Management Review, 23(3), 393-404.

Rudin, C., & Radin, J. (2019). Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From an Explainable AI Competition. Harvard Data Science Review, 1(2). https://doi.org/10.1162/99608f92.5a8a3a3d

Salah, M., Alhalbusi, H., Ismail, M. M., & Abdelfattah, F. (2023). Chatting with ChatGPT: decoding the mind of Chatbot users and unveiling the intricate connections between user perception, trust and stereotype perception on self-esteem and psychological well-being. Current Psychology. https://doi.org/10.1007/s12144-023-04989-0

Saunders, M. N., Dietz, G., & Thornhill, A. (2014). Trust and distrust: Polar opposites, or independent but co-existing? Human Relations, 67(6), 639–665. https://doi.org/10.1177/0018726713500831

Schadelbauer, L., Schlögl, S., & Groth, A. (2023). Linking Personality and Trust in Intelligent Virtual Assistants. Multimodal Technologies and Interaction, 7(6). https://doi.org/10.3390/mti7060054

Seymour, W., & Van Kleek, M. (2021). Exploring Interactions between Trust, Anthropomorphism, and Relationship Development in Voice Assistants. Proceedings of the ACM on Human-Computer Interaction, 5. https://doi.org/10.1145/3479515

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, 102551.

Simpson, J., & Weiner, E. (2021). Oxford English Dictionary. Clarendon Press.

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My Chatbot Companion - a Study of Human-Chatbot Relationships. International Journal of Human-Computer Studies, 149, 102601. https://doi.org/10.1016/j.ijhcs.2021.102601

Sundar, S. S. (2008). The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility. In M. J. Metzger & A. J. Flanagin (Eds.), Digital, media, youth, and credibility (pp. 72–100). Cambridge, MA: The MIT Press. https://doi.org/10.1162/dmal.9780262562324.073

Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., Zelaya, C. G., & 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 (pp. 272-283). https://doi.org/10.1145/3351095.3372834

Ueno, T., Sawa, Y., Kim, Y., Urakami, J., Oura, H., & Seaborn, K. (2022). Trust in Human-AI Interaction: Scoping Out Models, Measures, and Methods. In Extended Abstracts

of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22). Association for Computing Machinery, New York, NY, USA, Article 254, 1–7. https://doi-org/10.1145/3491101.3519772

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie, 64(7), 456–464. https://doi.org/10.1177/0706743719828977

Van Brummelen, J., Tabunshchyk, V., & Heng, T. (2021). Alexa, Can I Program You?": Student Perceptions of Conversational Artificial Intelligence before and after Programming Alexa. Proceedings of Interaction Design and Children, IDC 2021, 305–313. https://doi.org/10.1145/3459990.3460730

Vimalkumar, M., Sharma, S. K., Singh, J. B., & Dwivedi, Y. K. (2021). 'Okay google, what about my privacy?': User's privacy perceptions and acceptance of voice based digital assistants. Computers in Human Behavior, 120. https://doi.org/10.1016/j.chb.2021.106763

Wang, W., & Benbasat, I. (2005). Trust in and adoption of online recommendation agents. Journal of the Association for Information Systems, 6(3), 72–101.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., & Isaac, W. (2021). Ethical and social risks of harm from Language Models. arXiv. https://doi.org/10.48550/arXiv.2112.04359

Widyanto, L., & Griffiths, M. D. (2006). "Internet Addiction": A Critical Review. International Journal of Mental Health and Addiction, 4(1), 31–51. https://doi.org/10.1007/s11469-006-9009-9

Xie, T., Pentina, I., & Hancock, T. (2023). Friend, mentor, lover: Does chatbot engagement lead to psychological dependence? Journal of Service Management, 34(4), 806–828. https://doi.org/10.1108/JOSM-02-2022-0072

Yang, R., & Wibowo, S. (2022). User trust in artificial intelligence: A comprehensive conceptual framework. Electronic Markets, 32(4), 2053–2077. https://doi.org/10.1007/s12525-022-00592-6

Yuki, M., Maddux, W. W., Brewer, M. B., & Takemura, K. (2005). Cross-Cultural Differences in Relationship- and Group-Based Trust. Personality & Social Psychology Bulletin, 31(1), 48–62. https://doi.org/10.1177/0146167204271305

Zhang, Y., Gaggiano, J. D., Yongsatianchot, N., Suhaimi, N. M., Kim, M., Sun, Y., Griffin, J., & Parker, A. G. (2023). What Do We Mean When We Talk about Trust in Social Media? A Systematic Review. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3544548.3581019

Zierau, N., Engel, C., Söllner, M., & Leimeister, J. M. (2020). Trust in smart personal assistants: A systematic literature review and development of a research agenda. In International Conference on Wirtschaftsinformatik (WI), Potsdam, Germany.

Tables

Table 1

Overview of the 40 included articles

Overview		N	%
Year of publication	2013 – 2017	0	0
	2018	1	2.5
	2019	3	7.5
	2020	3	7.5
	2021	15	37.5
	2022	4	10
	2023	14	35
Geographical location	USA	11	27.5
	UK	3	7.5
	China	4	10
	Japan	1	2.5
	Thailand	1	2.5
	India	1	2.5
	Indonesia	1	2.5
	Singapore	1	2.5

	Taiwan	1	2.5
	Canada	1	2.5
	Brazil	1	2.5
	Iraq	1	2.5
	Multiple countries	6	15
	No information	7	17.5
Field of study*	Human-computer interaction	17	42.5
	Business/Marketing	12	30
	Information systems/technology/management	7	17.5
	Psychology	2	5
	Healthcare (including mental health)	2	5
Type of machine	Voice/Virtual assistants and smart speakers	29	72.5
	ChatGPT	3	7.5
	Text-based chatbots (e.g., Replika, Woebot)	5	12.5
	Researcher's developed chatbot	2	5
	Natural language generators (e.g., machine translators, chatbots, computer-assisted translation software)	1	2.5

Participants interacted with the machine during the study	Yes	7	17.5
	No	33	82.5
Subject of trust	Trust in the machine	36	90
	Trust in the manufacturer/producer	2	5
	Trust in the machine AND in the manufacturer/producer	2	5

Note. * Based on the article content or publication venue or first author's association.

Table 2

The 40 included articles' definitions of trust, operationalisations of trust, and methodology

Article number	Citation	Definition of trust	Operationalisation of trust	Methodology used
A1	Baughan et al. (2023)	A combination of confidence in a system as well as willingness to act on its provided recommendations (Madsen & Gregor, 2000)	Ability/Competence, Benevolence, Integrity, Willingness to use for future tasks	Interviews, cross-sectional survey, crowdsourcing
A2	Lau et al. (2018)	Did not define	NA	Diary study, interviews
A3	Van Brummelen et al. (2021)	Did not define	Trust/Trustworthy	Workshop
A4	Marikyan et al. (2022)	Did not define	Trust/Trustworthy, Reliable, Secure	Cross-sectional survey
A5	Pitardi & Marriott (2021)	A multidimensional concept that reflects perceptions of competence, integrity, and benevolence of another entity (Mayer et al., 1995)	Trust/Trustworthy, Ability/Competence, Integrity	Interviews, cross-sectional survey

A6	Malodia et al. (2023)	A user's confidence in the online system such that the user is willing to provide personal information and details that leave him or her vulnerable (Lee, 2005)	Trust/Trustworthy	Interviews, cross-sectional survey
A7	Müller et al. (2019)	Three trusting beliefs: competence, describing Alexa's ability to effectively perform in a specific domain, benevolence, referring to Alexa acting in the user's interest, and integrity, meaning that Alexa adheres to principles like promise-keeping and honesty (Wang & Benbasat, 2005)	Ability/Competence, Benevolence, Integrity	Cross-sectional survey
A8	Salah et al. (2023)	The extent to which a user is willing to depend on a technology and its outcomes (Mayer et al., 1995)	Ability/Competence, Benevolence, Perceived risk	Cross-sectional survey

Table 2. Continued

A9	Acikgoz et al. (2023)	A user's attitude of reliable expectation in response to the risk that their vulnerabilities will not be misused (Corritore et al., 2003)	Ability/Competence, Reliable, Safe	Cross-sectional survey
A10	Hasan et al. (2021)	A perception of risk that is dependent on another party's actions (Hu et al., 2010; Luo et al., 2010; Pavlou, 2003)	Ability/Competence, Benevolence, Integrity	Cross-sectional survey
A11	Oliveira et al. (2023)	A feeling of safety and comfort (Zierau et al., 2020; Pitardi and Marriott, 2021; Bawack et al., 2021)	Trust/Trustworthy, Ability/Competence, Benevolence, Integrity	Cross-sectional survey

A12	Hu et al. (2021)	A multifaceted construct comprising dimensions of integrity, competence, and benevolence, because the conversational AI possesses many human-like characteristics that traditional information systems do not have, such as human-like voice output and human-like understanding of users' voice input. (Lankton et al., 2015)	Ability/Competence, Benevolence, Integrity	Interviews, cross-sectional survey
A13	Seymour & Van Kleek (2021)	The extent to which a user is confident in, and willing to act on the basis of, the recommendations, actions, and decisions of an artificially intelligent decision aid (Madsen & Gregor, 2000)	Trust/Trustworthy	Cross-sectional survey
A14	Pal et al. (2021)	The extent to which users believe that using any technology will be reliable, credible, and safe (McKnight & Chervany, 2001)	Trust/Trustworthy, Willingness to use for future tasks, Reliable, Secure, Safe	Cross-sectional survey
A15	Bawack et al. (2021)	The belief that one party will not take advantage of the other's relative weakness but can rather depend on them to fulfill their commitments (Gefen et al., 2003)	Trust/Trustworthy, Reliable, Honest, Dependable	Cross-sectional survey
A16	Raiche et al. (2023)	Did not define	Benevolence, Credibility	Cross-sectional survey

Table 2. Continued

A17	Xie et al. (2023)	The willingness to be vulnerable to a social chatbot, based on the expectation that the social chatbot can offer emotional support and benefit users' wellbeing, without compromising their privacy and security (Mayer et al., 1995)	Benevolence, Reliable	Interviews, cross-sectional survey
A18	Chen & Park (2021)	Anticipations that the agent would possess the required aspects to be relied on (Komiak & Benbasat, 2006); faith reflecting emotional security (Rempel et al., 1985)	Ability/Competence, Integrity, Secure, Honest, Comfortable, Content	Cross-sectional survey
A19	Alimamy & Kuhail (2023)	The willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party (Mayer et al., 1995)	Trust/Trustworthy	Cross-sectional survey
A20	Poushneh (2021)	The degree to which consumers feel that their voice assistant is trustworthy and sincere and that they can rely on the assistant to perform tasks.	Trust/Trustworthy, Benevolence, Credibility	Experiment
A21	Prakash & Das (2020)	An individual's willingness to depend on another party because of the characteristics of the other party (McKnight et al., 2011; Li et al., 2008)	NA	User reviews

A22	Purwanto et al. (2020)	The expectation in the efficiency, reliability, and effectiveness of equipment and technical systems from the perspective of an individual who creates or a creator of a particular technology or material object (Ejdys, 2018)	Trust/Trustworthy, Secure, Helpfulness	Cross-sectional survey
A23	Choudhury & Shamszare (2023)	A user's willingness to take chances based on the recommendations made by this technology	Ability/Competence, Benevolence, Integrity, Reliable, Secure, Honest, Dependable, Credibility, Transparent	Cross-sectional survey

Table 2. Continued

A24	Baek & Kim (2023)	The belief that AI agents' recommendations and responses are reliable and credible (Shin, 2021)	Trust/Trustworthy, Credibility, Believable	Cross-sectional survey
A25	Schadelbauer et al. (2023)	Did not define	Reliable, Helpfulness, Structural assurance, Situation normality, Functionality	Cross-sectional survey
A26	Pfeuffer et al. (2019)	Did not define	Dependable	Experiment
A27	Skjuve et al. (2021)	Did not define	NA	Interviews

A28	Ashrafi & Easmin (2022)	A fiduciary relationship that ensures dependency, integrity and assurance to safeguard one's interests	Trust/Trustworthy, Ability/Competence, Integrity	Cross-sectional survey
A29	Vimalkumar et al. (2021)	Accumulated beliefs regarding the integrity, benevolence, and ability of the service provider to safeguard their interests (Mayer et al., 1995; McKnight et al., 2002)	Trust/Trustworthy, Reliable, Secure, Safe	Cross-sectional survey
A30	Cabrero-Daniel & Cabrero (2023)	Did not define	Trust/Trustworthy	Cross-sectional survey
A31	Pal et al. (2020)	A belief by the users where they can confide on certain entities to protect their personal information (Malhotra, 2004)	Trust/Trustworthy, Ability/Competence, Reliable, Secure	Cross-sectional survey
A32	Lee et al. (2021)	A psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another (Rousseau et al., 1998)	Trust/Trustworthy	Cross-sectional survey
A33	Hsu & Lee (2021)	Confidence in or reliance on some quality or attribute of a person or thing, or the truth of a statement (McAllister et al., 2006; Simpson & Weiner, 2021)	Trust/Trustworthy, Reliable, Secure, Humanlikeness	Cross-sectional survey

Table 2. Continued

A34	Hu et al. (2023)	Consumers' willingness to rely on their voice assistants and the level of confidence that voice assistants offer reliable and credible services. Perceived competence and benevolence of VA are two main backbones for developing trust in VA	Trust/Trustworthy, Reliable, Honest	Interviews, cross-sectional survey, longitudinal survey
A35	Acikgoz & Vega (2022)	Any person's willingness and belief to have confidence in another (Kumar, 1996)	Ability/Competence, Reliable, Safe	Cross-sectional survey
A36	Chandra et al. (2022)	The user's willingness to believe in the technology	Ability/Competence, Benevolence, Integrity	Cross-sectional survey
A37	Harrington & Egede (2023)	Did not define	NA	Diary study, interviews
A38	Liu et al. (2021)	The users' confidence in the reliability of smart homes to meet their needs and expectations	Trust/Trustworthy, Ability/Competence, Reliable, Controllable	Cross-sectional survey

A39	Clark et al. (2019)	Did not define	NA	Interviews
A40	Malodia et al. (2021)	Did not define	Trust/Trustworthy	Interviews, cross-sectional survey, content analysis

Table 3

Sample size information for the 40 included articles

		Sample Size			
Methods	Number of studies*	Mean	Median	Standard Deviation	[Min, Max]
Qualitative^	14	34	21	33	[12, 107]
Quantitative	36	333	290	189	[47, 732]

Note. * Some studies used multiple methods; ^ Excludes content analysis and user reviews because the involvement of human participants is unclear.

Table 4

Process of trust explored by the 40 articles

Process theme	Process categories	Article number
User-related	Demographics	A13, A16
	Knowledge	A3, A15, A26, A30, A38
	Personality	A7, A15, A16, A25
	Attitude	A13, A16, A25, A28, A35, A38
	Perception	A12, A19, A37
	Cognitive	A5, A6, A15, A28, A29, A33
Machine-related	Affective	A27, A28, A36
	Behavioural	A13, A18, A33, A34
	Practical	A5, A19, A24, A26, A27, A32, A36, A38, A39
Interaction-related		A1, A5, A6, A13, A18, A20, A24, A28, A32, A34, A38
Social-related		A6, A24, A28, A38
Context-related		A6, A13

Table 5

Outcomes of trust explored by the 40 articles

Outcome theme	Outcome categories	Article number
Affective	Attitude	A5, A9, A28, A33
	Satisfaction	A4, A11, A36
Relational	Brand loyalty	A10
	Customer experience	A11, A15
	Relationship quality	A17
Behavioural	Willingness to use	A1, A21, A36
	Continuance usage	A14, A19, A24, A31, A33
	Actual use/adoption	A23, A29
	Intention to use	A5, A6, A23, A28, A29, A30, A32
	Habit	A35
	Self-disclosure of personal information	A2, A9, A27, A31
	Engagement	A9, A17, A36
	Usage for functional task	A40
	Usage for information search	A40

	Engagement with work	A4
	Productivity at work	A4
	Recommendation to others	A11
Cognitive	Perceived performance	A22
	Performance expectancy	A29
	Perceived risks	A31
	Involvement	A19
	Self-brand connection	A19
Psychological	Self-esteem	A7
	Psychological well-being	A7
	Dependence	A17

Table 6

Future research suggestions

Category	Suggested research directions	Example questions
Methodological approaches	Longitudinal studies to explore the evolution of trust over time	What are the stages of trust development in AI-driven chatbots across long-term interactions?
		How does the evolution of trust differ between initial interactions and long-term use of AI-driven chatbots?
Mediating and moderating variables	Identification of the intermediary processes (mediators) and boundary conditions (moderators) affecting trust development	Which factors mediate the relationship between initial user perceptions and long-term trust in AI-driven chatbots?
		How do user characteristics moderate trust development in AI-driven chatbots?
Comparative studies	Comparative studies of trust dynamics across different cultural contexts and diverse populations	How does trust in AI-driven chatbots vary across different social, economic, or cultural groups?
		How do demographic factors influence users' trust in AI-driven chatbots?
	Comparative studies on how different chatbot types affect trust	What are trust-building factors for different types of AI chatbots?
	Comparison of real-time interactions and recall-based studies in trust research	How do real-time interactions with AI-driven chatbots differ

		from recall-based interactions in terms of trust development?
Psychological outcomes	Exploration of trust's impact on psychological outcomes	How does trust in AI-driven chatbots influence users' psychological well-being over time?
		What are the potential psychological risks and benefits of establishing trust in AI-driven chatbots for mental health support?
Trust dynamics	Investigation of users' trust in both the AI system and its manufacturer	What roles do users' perceptions of the manufacturer's play in shaping trust in AI-driven chatbots?
	Investigation of how distrust impacts user behavior	What factors contribute to the development of distrust in AI-driven chatbots?
		How does distrust affect users' interactions with AI-driven chatbots?

Figures

Figure 1

Flow diagram of included studies in which 40 articles were identified

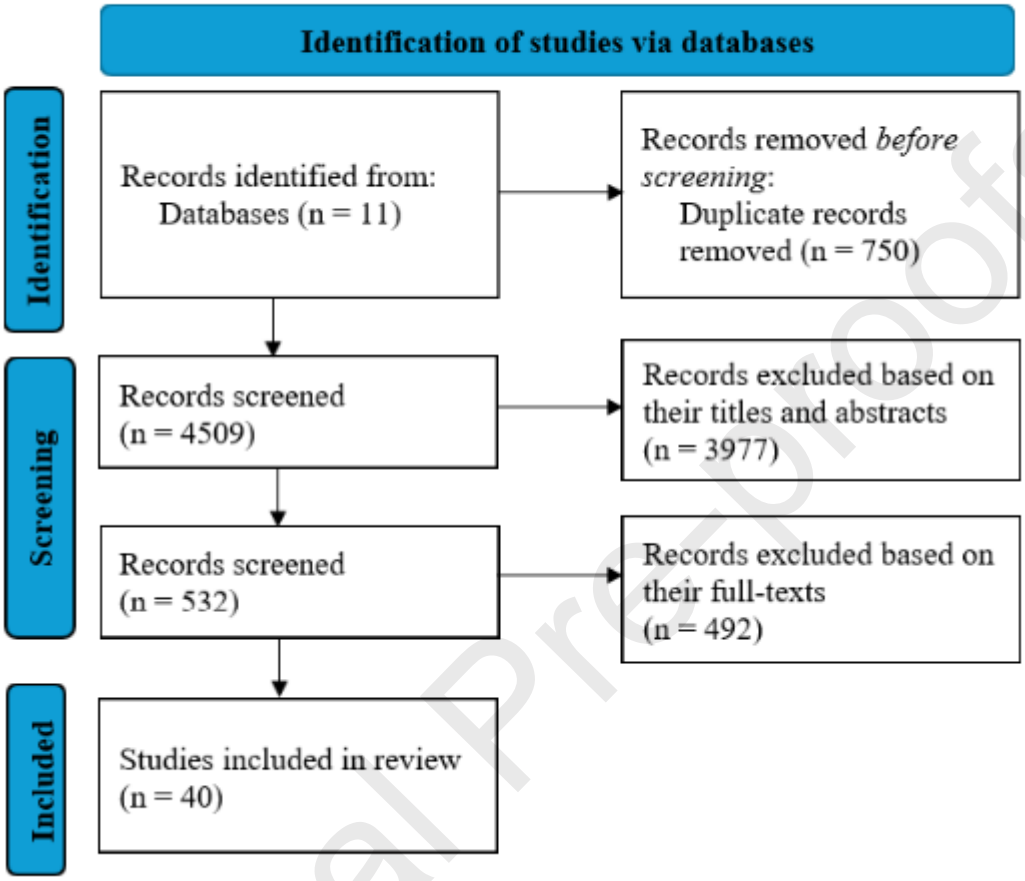


Figure 2

Visualisation of the distribution of trust conceptualisations

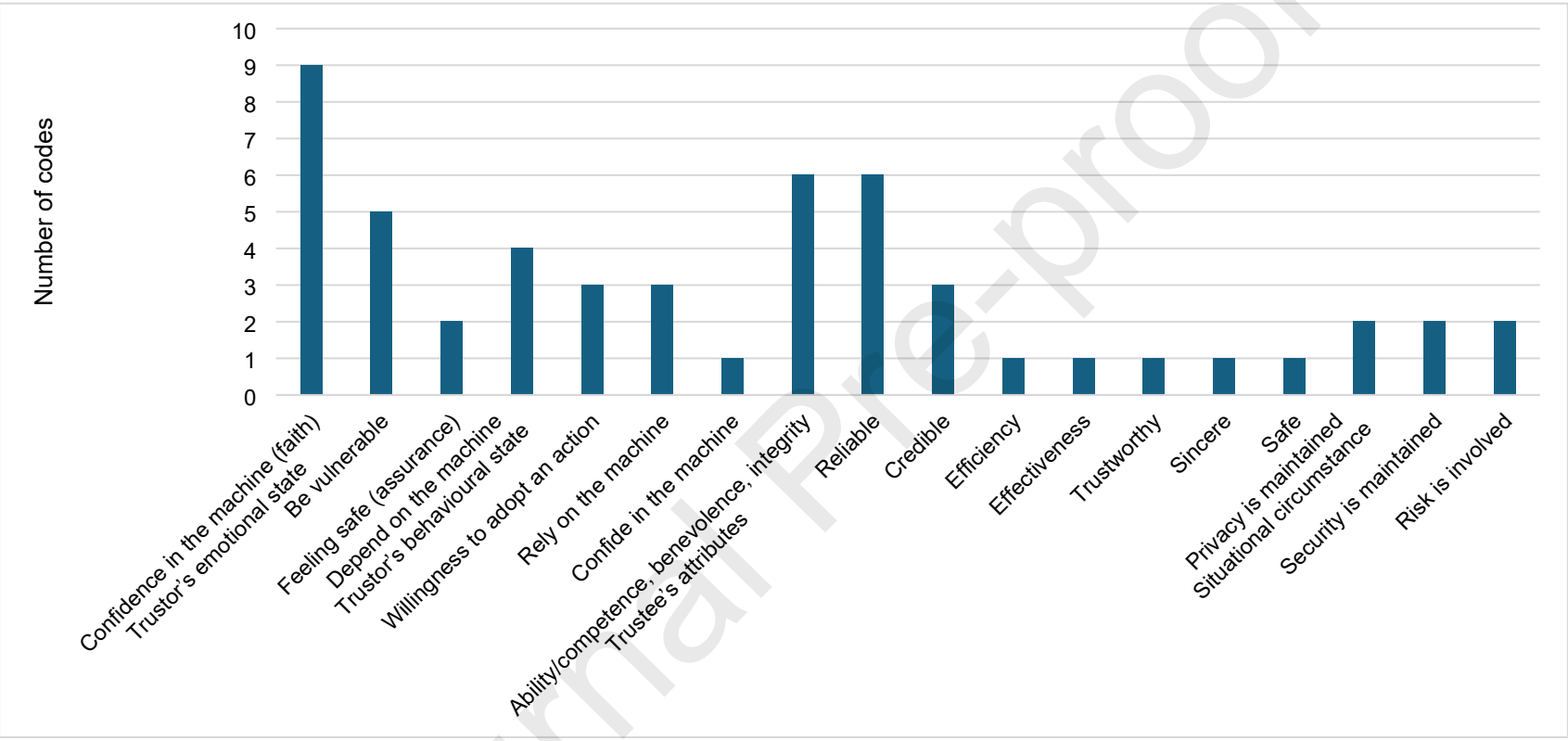


Figure 3

Visualisation of the distribution of trust operationalisations