Journal on Interactive Systems, 2026, 17:1
ISSN: 2763-7719 • doi: 10.5753/jis.2026.6787

Research Paper

Adoption of AI Chatbots by Users: A Systematic Literature Review on User Experience and Human Factors

Lívia Lutz [Pontifical Catholic University of Rio de Janeiro | *llutz@inf.puc-rio.br*]
Breno Lois Matos [Pontifical Catholic University of Rio de Janeiro | *brenoloismo3@gmail.com*]
Greis Francy M. Silva-Calpa [Pontifical Catholic University of Rio de Janeiro | *greis@inf.puc-rio.br*]

Department of Informatics, Pontifical Catholic University of Rio de Janeiro, Rua Marquês de São Vicente, 225, Gávea, 22451-900, Rio de Janeiro, Brazil.

Abstract. *Background:* Due to the increasing integration of chatbots into everyday digital systems, this study conducts a systematic literature review (SLR) to examine human–chatbot interactions. *Purpose:* The purpose of this study is to focus on the integration of artificial intelligence (AI) with human-computer interaction (HCI) principles to understand how chatbots are used across multiple contexts, how human factors and ethical considerations are addressed, and the impact of these interactions on UX and behaviour. *Methods:* The search query was applied across selected digital libraries to retrieve relevant studies. *Results:* From the reviewed studies, we found different types of chatbots used, the contexts of their use, how human and ethical factors are treated, user perceptions of the interactions, and broader implications for users' lives. The results reveal that chatbots used across different domains with various purposes make user interaction vary significantly by the context of use, showing both positive and negative impacts and raising ethical concerns. *Conclusion:* Regarding open gaps, aspects such as the long-term impact of chatbots, the use of AI tools by people with physical disabilities, and non-generic UI/UX aspects were missing.

Keywords: AI Chatbots, Users, User Experience, Human Factors, Artificial Intelligence

Edited by: Milene S. Silveira | **Received:** 31 October 2025 • **Accepted:** 02 August 2026 • **Published:** 23 August 2026

1 Introduction

Artificial intelligence (AI) refers to the capability of computational systems to perform tasks typically associated with human intelligence, such as learning, reasoning, problem solving, perception, and decision making [Wikipedia contributors, 2025]. Chatbots are computer programs designed to simulate human conversation through text or speech by using AI techniques to understand user input and provide relevant responses [Wikipedia contributors, 2024]. As AI-driven applications, chatbots combine different computational approaches to interpret and generate language in ways that resemble natural human interaction. Natural Language Processing (NLP), a subfield of AI, is concerned with enabling computers to learn, understand, and process human language [Hirschberg and Manning, 2015].

Allowing chatbots to interpret the structure and meaning of user input. In addition, many chatbots rely on Machine Learning (ML) models that improve over time through user interactions. These models can adapt their responses based on previous conversations, user preferences, and feedback, thereby enhancing the overall user experience (UX) [Oracle, 2024].

In this context, chatbots represent a distinct category of AI-based interactive systems. They are distinguished by their language-based interaction model, in which natural language serves as the primary interface between the user and the system. Such interaction is generally structured as turn-taking dialogue and relies on context-aware processing, which shapes user engagement and mediates perceptions of system responsiveness and adaptivity [Weber and Ludwig, 2020]. As human–chatbot engagement continues to grow and these systems become increasingly embedded in everyday digital environments, it becomes essential to examine how such interactions are designed and experienced by users [Weber and Ludwig, 2020; Kothari and Noel Newton Fernando, 2024; Virvou and Tsihrintzis, 2024].

Over the last decade, mobile messaging tools like WhatsApp, Telegram, and Facebook Messenger have grown significantly, with chatbots incorporated as part of this technological ecosystem [Merelo Guervos *et al.*, 2023]. As a result, these platforms have contributed to making chatbot usage a natural part of everyday digital interactions. With the evolution and integration of AI, they have transformed from simple tools into constant companions capable of fulfilling users' daily tasks and communication needs [Skjuve *et al.*, 2021]. As pointed out by [Adamopoulou and Moussiades, 2020a], this widespread adoption has reshaped users' expectations regarding the availability, responsiveness, and intelligence of digital systems, reinforcing the importance of emotional awareness, contextual relevance, and ethical design in chatbot development. [Silva and Canedo, 2024] highlights that conversational design must be approached as a distinct human-centred discipline, in which textual, visual, and interactive choices directly enhance or undermine user satisfaction, engagement, and trust. This reinforces the need to integrate human factors such as transparency, naturalness, and emotional resonance to prevent interaction breakdowns and meet users' nuanced expectations.

Diverse studies already discuss the impact of chatbots [Rapp *et al.*, 2021; Adamopoulou and Moussiades, 2020b; Luo *et al.*, 2022; Park *et al.*, 2022; Caldarini *et al.*, 2022; Aumüller *et al.*, 2024], most of which focus on specific aspects such as consumer satisfaction, design considerations, and harm to graduate students. Articles that discuss the use of

conversational agents (CAs) in customer service tend to have a more commercial perspective and aim to determine how to shape chatbots to gain users' trust and satisfaction [Luo *et al*., 2022; Min *et al*., 2021; Samonte and Lising, 2024; Putra Antonio *et al*., 2023]. In contrast, studies in domains such as mental health or student learning tend to be goal-orientated and context-specific, often focusing on usability and explainability within particular applications [Meng *et al*., 2023; Park *et al*., 2024; Siddals *et al*., 2024; Maples *et al*., 2024; Pozzi and De Proost, 2024; Wicaksono *et al*., 2024; Yeh *et al*., 2025; Rodríguez-Ruiz *et al*., 2025].

Although there is a growing body of studies on the chatbot–user relationship, existing research has concentrated on specific or isolated aspects, such as adoption or trust [Ng and Zhang, 2025; Mehrotra *et al*., 2024; Alharbi *et al*., 2025], user experience [Rapp *et al*., 2021; Mariano *et al*., 2024; Guerino and Valentim, 2020], or ethical and privacy concerns [Silva and Canedo, 2025; May and Denecke, 2021; Gumusel, 2025]. Fewer studies offer a systematic and integrated synthesis across these dimensions. As a result, the interrelations between chatbot types and application contexts, interaction characteristics and user experience, reported impacts on users' behaviour and lives, and the ways human and ethical aspects are addressed remain insufficiently articulated in the literature, limiting a more comprehensive understanding of human–AI chatbot interaction.

Therefore, to address this gap, we conduct a systematic literature review (SLR) on human-chatbot interaction, using an integrated perspective that jointly examines chatbot types and application contexts, interaction characteristics, user experience, reported impacts on users' lives, and how human factors and ethical aspects are addressed. Through this SLR, we aim to answer a set of research questions (RQ) designed to map the types of chatbots and their contexts of use, clarify the purposes of the studies, analyse how interactions are conducted and perceived by users, and examine reported impacts on users' lives. The review also identifies how user experience is assessed and how human factors and ethical aspects are addressed in the literature.

In summary, the main research question guiding this review is: How do studies in the HCI literature describe and evaluate human–chatbot interactions, and what impacts on user experience, behaviour, human factors, and ethical considerations are reported across application contexts?

Results show a predominance of chatbots, especially ChatGPT, used across various domains for different research purposes. The user interaction varied significantly by context, and users generally appreciated convenience, responsiveness, emotional design, and avatar representation; however, failures or a lack of empathy often led to reduced trust. Additionally, the studies demonstrated that chatbots can positively impact learning, decision-making, and emotional well-being. Ethical concerns are central across the majority of the reviewed studies, highlighting trust, privacy, bias, epistemic injustice, and responsible design.

This article is structured as follows. The next section presents the background and discusses related work relevant to this review. In the Systematic Literature Review section, we outline the methodological process, including inclusion and exclusion criteria, information sources, search strategy, and review process. We then present the results, detailing the study selection and characteristics, followed by a discussion of the findings.

2 Background

ELIZA was one of the earliest chatbots, developed using rule-based patterns and word-substitution techniques without any form of ML [Limna *et al*., 2023]. ELIZA simulated the operation of a psychotherapist but could only generate answers to a specific domain of topics and could not hold long conversations, learn, or understand the context of the discussion. Over time, chatbot technology has transitioned from pattern-matching approaches to more sophisticated systems powered by ML and NLP. Today, chatbots are able to understand user intent, manage context, and even detect emotional tone. However, several challenges remain regarding the use and adoption of these tools, particularly in ensuring safety, ethical compliance, and user-centred design [Adamopoulou and Moussiades, 2020a].

Chatbots have become very popular for several reasons. First, their implementation is platform-agnostic and requires minimal configuration [Adamopoulou and Moussiades, 2020b]. They can also interpret conversation context and answer questions 24/7, making chatbots a user-friendly tool that can assist humans with virtually any type of knowledge [Iberdrola, 2025]. Their role has expanded to include not only functional interactions but also persistent engagement with users through notifications and integration with messaging platforms [Adamopoulou and Moussiades, 2020b]. This technology can now be found integrated into various devices and digital services, such as virtual assistants like Google Assistant, Apple's Siri, and Amazon's Alexa, gaining adoption among users across diverse social and demographic backgrounds.

Their integration into everyday digital systems has influenced how users interact with technology, particularly in terms of how they perceive these systems, the emotional impact of this technology, and their expectations and difficulties in using them. These aspects highlight the importance of evaluating technological advancements in alignment with human factors and HCI principles. This evaluation is key to ensuring that these technologies enhance UX while maintaining trust, clarity, and fairness.

According to [Belda-Medina and Calvo-Ferrer, 2022], speech technologies (R&S) were the main reason for successful human-like communication between a bot and a person. However, the similarity between a chatbot's behaviour and that of a human being can present some risks.

Recent studies have highlighted concerns regarding the broader societal and economic impacts of OpenAI's ChatGPT, which bring issues such as job displacement, the digital literacy gap, and AI-induced anxiety [Dempere *et al*., 2023]. Additionally, a study by [Kooli, 2023] mentions the importance of acting in a responsible and transparent manner, stating that it could even destroy the aim of education as a mechanism for knowledge and capacity-building processes.

Given the rapid evolution and widespread adoption of chatbots, many research studies have attempted to better understand their design, implementation, and impact on users.

The following section examines these studies in detail, highlighting their main contributions, objectives, and limitations in order to state our objectives and contributions with this review.

3 Related Work

There are similar studies in the literature related to our research, as they also examine human-chatbot interaction, each focusing on different aspects than our SLR. Overall, secondary research studies can be grouped into four categories based on their main research objectives: Technical and architectural aspects [Park *et al*., 2022; Caldarini *et al*., 2022; Adamopoulou and Moussiades, 2020a], user interaction and UX [Rapp *et al*., 2021], domain-specific applications [Luo *et al*., 2022], and evaluation methods and metrics [Maroengsit *et al*., 2019]. While these studies provide valuable insights, they address these dimensions in isolation rather than offering an integrated perspective on human–chatbot interaction.

Regarding studies focusing on Technical and architectural aspects, one paper [Park *et al*., 2022] stated the need to review the technology, with a focus on recent advancements and core research. Therefore, it introduces five different chatbot technologies: NLP, pattern matching, semantic web, data mining, and context-aware computing, while also introducing the latest technology for chatbot researchers to recognise the current situation and channel it in the right direction. A different study [Caldarini *et al*., 2022] reviewed recent advances in chatbots where AI and NLP are used, highlighting the main challenges and limitations of current work and making recommendations for future research investigations. Although these studies addressed the issue of chatbot technology and its limitations, they did not consider any UX, ethical, or human aspects of the user-chatbot interaction, which are the aims of this SLR. Another study [Adamopoulou and Moussiades, 2020b] discusses the motivations that drive the use of chatbots, clarifying their usefulness in a variety of areas, highlighting the impact of social stereotypes on chatbots' design, and presenting the general architecture of modern chatbots while mentioning the main platforms for their creation.

For User interaction and UX, one article [Rapp *et al*., 2021] discusses how users interact with text-based chatbots, mapping the relevant themes, describing how people experience the chatbot in terms of satisfaction, engagement, and trust, and addressing whether and why they accept and use this technology, how they are emotionally involved, what kinds of downsides can be observed in human-chatbot conversations, and how the chatbot is perceived in terms of its humanness. Although it explored many aspects of interaction, it did not examine how ethical aspects were treated in much detail, as we explored in this SLR.

For Domain-specific applications, the study by [Luo *et al*., 2022] critically analysed various computational approaches used to develop state-of-the-art chatbots, also reviewing the usability and applications of chatbots within the business sector. The authors specifically focused on the business sector, while our SLR concentrates on understanding how chatbots are used across multiple contexts.

Lastly, for Evaluation methods and metrics, the survey by [Maroengsit *et al*., 2019] presents a classification of chatbot evaluation methods, analysing them in relation to different chatbot types and identifying three primary evaluation dimensions: content evaluation, user satisfaction, and chat function. More recently, [Rocha *et al*., 2025] conducted a systematic literature mapping specifically on UX evaluation technologies for chatbots, cataloguing the methods and instruments used to assess user experience in this domain. Although the user perceptions of the interactions were taken into account, aspects such as the different types of chatbots used, the contexts of their use, how human and ethical factors are treated, and the broader implications on users' lives were not considered; these are the aspects we aim to understand with this SLR.

Regardless of the many studies examining chatbots, few directly explore their relationship with users and the resulting impact, especially from an ethical and behavioural standpoint. Most existing research centres on specific domains like customer service, mental health, or education, often with commercial or practical goals. However, none of these studies provide a comprehensive synthesis that simultaneously considers UX, ethical issues, and human factors across multiple contexts. Recognising this gap, we conducted a more in-depth investigation into aspects that were not explored.

Observing this gap, this study proposes a review that builds upon existing studies of CAs and chatbots by offering a human-centred analysis of chatbot use. Our study specifically examines the relationship between humans and chatbots, and we seek to explore how such interactions affect UX and user behaviour, as well as how human, ethical, and emotional factors are considered in the current research landscape. We integrate UX, human factors, and ethical dimensions into a single synthesis focused on chatbots across multiple contexts.

In this review, we adopt the term chatbots to refer broadly to conversational systems that interact with users primarily through natural language interfaces. The reviewed studies encompass a range of technological implementations. Among them, seven studies examined Large Language Model (LLM)-based chatbots, most commonly ChatGPT, while two focused on supervised machine learning–based systems. Several studies did not clearly specify the underlying chatbot technology, leaving the chatbot type indistinct. Although the technologies varied across the literature, technological architecture was not used as an inclusion or exclusion criterion in this SLR, as the focus of the review is on human–chatbot interaction and its reported impacts rather than on specific AI implementation approaches.

Table 1 shows a comparison of existing secondary studies, highlighting their main objectives as well as their limitations, and explicitly demonstrating how our review differs from prior work.

4 Systematic Literature Review (SLR)

This section outlines the SLR method, following the guidelines established by [Kitchenham *et al*., 2009] and cross-referenced with the PRISMA 2020 expanded checklist to ensure comprehensive reporting [Page *et al*., 2021a,b]. The steps in the SLR method are documented below.

Table 1. Comparison of existing secondary studies on chatbots and the contribution of this SLR.

Category	Definition and Limitations	Studies
Technical and Architectural Approaches	Studies that focus on chatbot architectures, NLP techniques, and system design. These works emphasise implementation and performance, with limited attention to user experience, ethical concerns, or behavioural impact.	[Adamopoulou and Moussiades, 2020b]; [Park *et al*., 2022]; [Caldarini *et al*., 2022]
User Interaction and Experience	Studies that explore user perceptions, engagement, trust, and emotional responses to chatbots. However, they tend to focus on interaction outcomes without deeply addressing ethical implications or broader human factors across contexts.	[Rapp *et al*., 2021]
Domain-Specific Applications	Studies that analyse chatbot use within specific domains (e.g., business), focusing on usability and efficiency in particular contexts. These works lack a cross-domain perspective and do not provide a holistic understanding of human–chatbot interaction.	[Luo *et al*., 2022]
Evaluation Methods and Metrics	Studies that classify or propose evaluation methods for chatbots, focusing on metrics such as content quality and user satisfaction. They do not consider broader ethical, behavioural, or human-centred implications of chatbot use.	[Maroengsit *et al*., 2019]; [Rocha *et al*., 2025]
Adoption of AI Chatbots by Users: A Systematic Literature Review on User Experience and Human Factors (Our study)	Our study addresses the limitations identified above by providing a comprehensive synthesis that integrates UX, human factors, and ethical considerations across multiple domains. It offers a cross-context analysis of how chatbot interactions influence user behaviour, perception, and societal impact.	Our study

4.1 Research Questions

The review was guided by the following Research Questions (RQ):

- RQ1: What types of chatbots have been analysed and in which application contexts?
- RQ2: What research aims (i.e., the main purposes of the studies) are addressed in the analysed studies?
- RQ3: How are human–chatbot interactions characterised and evaluated in the literature, and what user perceptions and experiences are reported across different application contexts?
- RQ4: What impacts of chatbot use on users' behaviour and everyday life are reported? RQ4.1: Within these impacts, what evidence regarding users with physical or cognitive difficulties is reported?
- RQ5: How are human factors and ethical considerations addressed in the analysed studies?

With respect to RQ1 and RQ2, these questions provide the contextual basis for the review and support meaningful comparisons across studies. RQ1 identifies the types of chatbots and their application contexts, while RQ2 examines the main purposes of each study. Understanding these elements helps explain variations in reported outcomes and evaluation criteria across the literature.

On this basis, RQ3 examines how human–chatbot interactions are described in the reviewed studies and how user experience (UX) is evaluated and reported through a thematic synthesis of interaction contexts, UX evaluation approaches, and user perceptions. RQ4 focuses on the reported impacts of chatbot use on users' behaviour and everyday lives, contributing to an understanding of the broader implications of chatbot adoption in different contexts. RQ4.1 complements RQ4 by applying an inclusion lens, examining whether everyday-life impacts are explicitly reported for users with physical or cognitive difficulties and which inclusion or support-related considerations are discussed. Finally, RQ5 explores how human factors and ethical considerations are addressed in the analysed studies, enabling a comparative synthesis of these aspects within the research domain.

Together, these questions move from contextual description to analytical synthesis, supporting a comprehensive understanding of interaction characteristics, user experience assessment, reported impacts, accessibility considerations, and the treatment of human and ethical aspects in the literature.

4.2 Review Protocol

A review protocol was defined prior to the execution of the search process, specifying the research questions, search strategy, inclusion and exclusion criteria, and planned data extraction fields. The protocol was internally documented and used as a guiding framework throughout the review to ensure consistency and traceability of decisions. A pilot search was conducted to refine the search string and verify the adequacy of the inclusion and exclusion criteria before applying them across all databases.

The protocol was not externally registered but was internally documented and consistently followed throughout the review.

4.3 Search Process

The review was conducted using the ACM Digital Library, IEEE Xplore, SpringerLink, ScienceDirect, and Web of Science, which were selected to cover studies in Computing, HCI, AI-based interactive systems, and related interdisciplinary areas. During the screening process, Web of Science contributed only one study to the final corpus after duplicate removal and eligibility assessment, indicating a high degree of overlap with the other selected sources. This reinforced the expected overlap among broad indexing databases. Although we recognise that Scopus is also a relevant source for identifying peer-reviewed studies in these fields, it was not included because adding another broad indexing database would likely increase duplicate records and screening workload. We acknowledge that some relevant studies indexed only in Scopus may, therefore, have been missed. This is considered a limitation of the review and a potential threat to

external validity.

The search dates for each database are:

- ACM Digital Library: January 15, 2026
- IEEE Xplore: January 16, 2026
- SpringerLink: January 17, 2026
- Web of Science: February 5, 2026
- ScienceDirect: February 5, 2026

To ensure consistency and comprehensive coverage across these platforms, we developed a standardised search string: (Artificial Intelligence OR AI) AND (AI-chatbot OR chatbot*) AND Interaction AND user*.

The final search strategy was defined through an iterative process involving three pilot tests. We included broader terms such as “conversational agents" and “dialogue systems" in the search string. However, during the pilot phase, these terms retrieved a large volume of studies that were not within the proposed context and did not align with our focus on modern AI-driven human-chatbot interaction. In particular, they introduced excessive noise by retrieving studies unrelated to the intended scope of human-chatbot interaction in AI-driven contexts. They also lacked the specificity needed to capture the contemporary intersection of AI, UX, and human factors. For example, these terms retrieved studies on general dialogue systems, voice assistants, NLP architectures, and conversational systems without empirical user evaluation, which were outside the scope of this review.

We found that the term "chatbot" (and its variations) functioned as the most effective overarching term for the contemporary period from 2020 to 2026, retrieving studies that specifically discuss the intersection between AI, UX, and human factors. Restricting the search string to this term and its variants allowed us to maintain a focused and highly relevant set of studies, ensuring that our analysis was conducted within the specific context of chatbots.

Although a publication time frame was applied, the issue was not limited to the retrieval of older studies but rather to the lack of specificity of these broader terms, which reduced the precision of the search results. Consequently, the term “chatbot” (and its variations) was retained as the primary search term, as it more effectively captured studies within the intended scope, particularly those addressing the intersection of AI, UX, and human factors in recent research (2020-2026).

This refinement of the search string improved the relevance and focus of the retrieved studies, ensuring alignment with the objectives of the review while maintaining a manageable and methodologically rigorous screening process.

Importantly, the underlying syntax for constructing search queries was consistent across all five digital libraries, allowing us to apply the exact same refined query to each. The search query is: ("Artificial Intelligence" OR "AI") AND ("AI-chatbot" OR chatbot*) AND "Interaction" AND user*. Minor syntax adaptations were applied when required by specific database interfaces while preserving the semantic structure of the query. The use of quotation marks around "Artificial Intelligence" and "Interaction" ensured these were treated as exact phrases, while the asterisk (*) in chatbot* and user* allowed for variations like "chatbots" or "users," respectively. The Boolean operators AND and OR were used to combine these terms effectively, targeting articles that discussed artificial intelligence or AI, specifically in relation to chatbots or AI-chatbots and their interaction with users.

Our search was based on titles; we did not use full text search. The filters applied pertained to the time frame and studies in English, and we excluded the filters for surveys, reviews, book chapters, and posters.

The screening process was conducted independently by two reviewers at the title and abstract stages. Each reviewer applied the inclusion and exclusion criteria to the retrieved records. Disagreements were discussed until consensus was reached. When consensus could not be achieved, a third reviewer was consulted to make the final decision.

Because individual pre-consensus screening decisions were not recorded in a structured format, it was not possible to calculate an inter-rater reliability measure, such as Cohen’s kappa. This is acknowledged as a limitation of the screening process. Nevertheless, independent screening, consensus discussion, and third-reviewer arbitration were used to reduce selection bias.

No automated inclusion decisions were performed; screening tools such as Zotero[1], Google Sheets[2], and Rayyan[3] were used to support the screening process and apply the selection criteria.

Rayyan was used as a semi-automatic screening tool to organise records, apply the inclusion and exclusion criteria, and filter articles by title and abstract, facilitating the filtering, tagging, and systematic application of criteria rather than making automatic inclusion decisions. The ultimate inclusion/exclusion judgement remained human-driven, being performed manually by the authors based on full-text assessment where required. This approach mitigates the risk that automated filtering could bias study selection, as Rayyan was not used to replace human screening but to support transparent and consistent decision-making.

4.4 Inclusion and Exclusion Criteria

Inclusion and exclusion criteria were defined to guide the systematic selection of studies aligned with the research questions.

Studies were included if they met the following criteria:

- The study investigates human–chatbot interaction and reports empirical user evaluation.
- The full text of the article is available.
- Articles published in the last 6 years (Jan 2020 - Jan 2026)
- The publication is peer-reviewed (journal or conference paper).

Articles based on the following criteria were excluded:

- Duplicate articles across databases.
- Articles not in English.
- Studies focusing solely on technical AI development without human interaction or user evaluation.
- Secondary studies (e.g., literature reviews or surveys).

[1] `https://www.zotero.org`. Accessed on 15 June 2026

[2] `https://workspace.google.com/products/sheets/`. Accessed on 15 June 2026

[3] `https://rayyan.ai/`. Accessed on 15 June 2026

- Book chapters, posters, editorials, or non-peer-reviewed publications.

The time frame adopted in this study (Jan/2020–Jan/2026) was selected to capture recent developments in human–chatbot interaction, particularly related to UX, behavioural responses, and ethical considerations. This period reflects the rapid evolution of conversational AI, during which the deployment of more advanced chatbot systems has greatly influenced the way users interact with these technologies. As a result, studies published within this time frame are more likely to address contemporary interaction patterns and user-related outcomes.

While advances such as LLM-based chatbots, including systems like ChatGPT, have contributed to this shift, the motivation for this time frame is to examine human-centred aspects of chatbot use rather than the underlying technological architecture. We excluded studies published prior to this period, as they focused more on earlier generations of CAs with different interaction characteristics from the current systems. Therefore, this restriction was applied to keep the review aligned with its objective of analysing recent developments in human–chatbot interaction from a human-centred perspective.

4.5 Quality assessment

A formal methodological quality scoring system was not applied to rank or exclude studies. Instead, quality control was ensured through predefined inclusion criteria, including peer-reviewed publication status and the requirement for empirical user evaluation. Studies focusing solely on technical AI development without human interaction were excluded to maintain alignment with the review scope. During the full-text assessment, methodological clarity, transparency of research design, and the presence of reported empirical evidence were considered to ensure the analytical robustness of the included studies. Given the exploratory and integrative objective of this review, studies were included based on their relevance to the research questions and empirical grounding rather than on numerical quality scoring.

4.6 Data Extraction

After completing the screening process, the remaining articles were analysed in relation to each RQ. The studies were then grouped based on common keywords and thematic subjects to facilitate synthesis.

A structured data extraction form was developed using Google Sheets [4] to collect the extracted data from each article reviewed. The data included the chatbot type and application context (RQ1), study objectives (RQ2), interaction design and UX evaluation methods (RQ3), reported behavioural and life impacts (RQ4, RQ4.1), and human factors as well as ethical considerations (RQ5).

Each study was coded independently by two reviewers to ensure consistency. Discrepancies were resolved through discussion.

[4] `https://workspace.google.com/products/sheets/`. Accessed on 15 June 2026

4.7 Data Synthesis

Quantitative data were summarised using frequency counts to describe distributions across chatbot types, application contexts, and reported outcomes. Qualitative data were analysed using a coding approach. A primarily deductive coding scheme was defined in advance based on the research questions (RQs), with initial analytical categories derived directly from them. Emerging subthemes identified across studies were incorporated inductively during the analysis when patterns were observed beyond the predefined categories. The synthesis aimed to identify patterns, relationships, gaps, and trends in how human–chatbot interactions and their reported impacts are addressed in the literature.

The coding process was conducted independently by two reviewers. Studies were assigned to categories according to the predefined coding scheme. Discrepancies in coding were discussed until consensus was reached. This procedure aimed to enhance analytical consistency and reduce subjective bias in the synthesis process.

Figure 1 illustrates each stage of the review process, including the number of studies retrieved, screened, excluded (with reasons), and ultimately included in the final analysis. At the end of this paper, Tables 9 - 12 show a summary of all 56 reviewed articles after the screening process, identified as S1, S2, ..., S56. The majority of the records excluded during the title and abstract screening phase were removed because they did not include UX or ethical considerations of the examined chatbots. One article with the title “Impact of artificial intelligence on human loss in decision making, laziness, and safety in education" was excluded after the eligibility assessment because it was retracted from its journal.

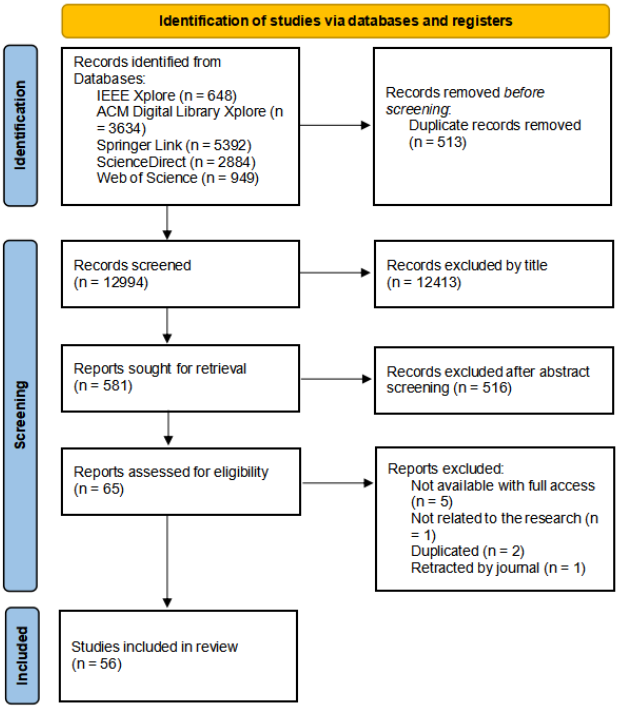

Figure 1. Flow diagram adapted from the PRISMA statement [Page *et al.*, 2021b] [Page *et al.*, 2021a]

Figure 2 shows the number of articles selected per database/digital library, with sixteen from *IEEE Xplore*, sixteen from *SpringerLink*, thirteen articles from *The ACM Dig-*

ital Library, ten articles from *ScienceDirect*, and one from *Web of Science*. Of the 56 articles, twenty-nine were published in conference proceedings and twenty-seven were published in journals. Most of the studies were published in 2024, as shown in Figure 3.

Table 2 presents a structured mapping of analytical categories, their definitions, and the corresponding studies (S1,S2,S3,...) assigned to each category.

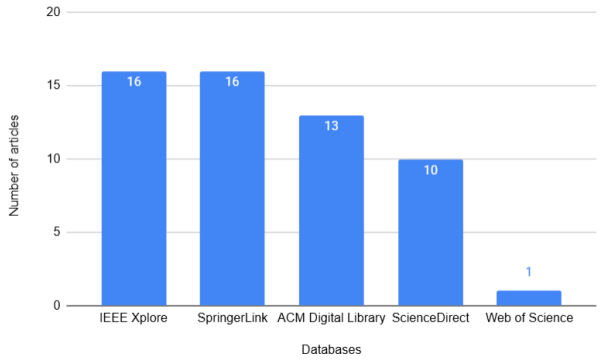

Figure 2. Number of included articles per database

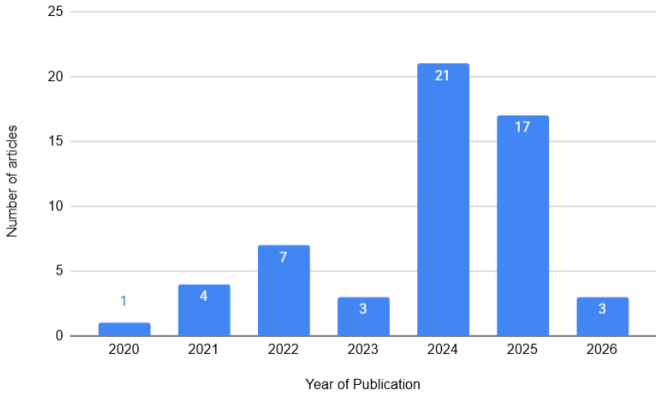

Figure 3. Number of included articles per publication year

5 Results

In this section, we present the results found after answering all five RQs following the SLR procedure.

5.1 RQ1: What types of chatbots have been analysed and in which application contexts?

Chatbots are categorised by their operational scope (generic, cross-domain or open-domain, domain-specific) and function (conversational or task-based) [Adamopoulou and Moussiades, 2020a]. Within the operational scope, chatbots can be classified as Generic chatbots, such as ChatGPT and Amazon's Alexa. Another category is Cross or Open-Domain chatbots, which operate in more than one domain and are represented by these general-purpose systems. On the other hand, Domain-Specific chatbots focus on a particular field. Within the operational scope, chatbots can be classified as Conversational or Task-Based chatbots. The first is for natural conversation with users, while the second is for handling different functions, helping users respond appropriately or request information, which is suitable for a commercial context [Adamopoulou and Moussiades, 2020a]. Our review did not focus on any specific type of chatbot, but we observed a predominance of generic and conversational chatbots, as they are the most integrated into people's daily lives and are found in various contexts, including psychological support, learning environments, and general assistance.

From the reviewed studies, we found different types of chatbots being used in various contexts. Table 3 provides an overview of the types of chatbots used in each study, as well as the contexts in which they were utilised. Figure 4 consolidates the number of articles per subject, indicating which articles mentioned chatbots used in behavioural, consumer, education, or mental health contexts.

The classifications shown in Figure 4 were not predefined from a fixed taxonomy in prior literature. Instead, they were derived inductively through iterative coding based on the context of use in which each study evaluated its chatbot. Studies were coded at the descriptive level by identifying the chatbot's application context.

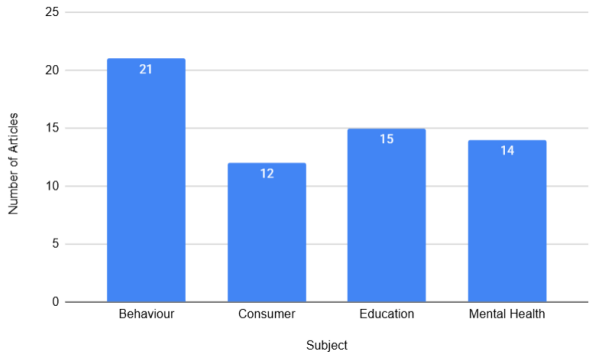

Figure 4. Number of articles per subject

In terms of types, AI-driven systems such as ChatGPT and general chatbots dominate the research landscape. As shown in Table 3, fourteen studies [S1, S19, S22, S23, S28, S30, S32, S40, S45, S48, S50, S51, S52, S53] were carried out using OpenAI's implementation, which, being the most used among all studies, is only behind those conducted with general chatbots, without analysing a specific model (19 in total). The high number of studies on OpenAI ChatGPT is a reflection of its global popularity, with a user base of over 180.5 million as of August 2023 [Deng *et al.*, 2023]. These AI tools leverage Large Language Models (LLMs) to generate natural, human-like responses and are widely used in educational, psychological help, and decision-making contexts. Alongside these, other AI tools integrate NLP and contextual reasoning techniques to support dynamic, multipurpose interactions.

Four articles [S1, S2, S4, S8] did not delve into any context beyond AI-HCI; of these four studies, only one examined a type of chatbot [S1]: ChatGPT.

When examining the usage context, we found a strong alignment between chatbot types and domains. Educational chatbots, often powered by LLMs like ChatGPT, are tailored to promote learning outcomes and critical thinking. Other categories, such as customer service chatbots, tend to focus on simulating human dialogues, often with goal-orientated or

Table 2. Context-based categories, definitions, and corresponding studies.

Category	Definition	Studies
Behaviour	Studies that examine how chatbots influence users' behaviour, decision-making, ethical actions or behavioural outcomes.	S1 [Fabio *et al*., 2024], S2 [Akhtar *et al*., 2024], S3 [Khurana *et al*., 2021], S4 [Joshi and Bengler, 2024], S5 [Meng *et al*., 2024], S6 [Mubassira *et al*., 2025], S7 [Zimmerman *et al*., 2024], S8 [Weber and Ludwig, 2020], S9 [Min *et al*., 2021], S10 [Aumüller *et al*., 2024], S11 [Kim *et al*., 2023], S19 [Rodríguez-Ruiz *et al*., 2025], S38 [Wong *et al*., 2021], S42 [Genç *et al*., 2025], S43 [Shahid *et al*., 2025], S45 [Wang *et al*., 2025], S46 [He *et al*., 2025], S49 [Kang *et al*., 2025], S50 [Mayerhofer *et al*., 2025], S55 [Tsumura and Yamada, 2025], S56 [Thalhammer, 2025]
Consumer	Studies that evaluate chatbots in consumer contexts such as customer service, e-commerce, online shopping, service satisfaction and customer experience.	S9 [Min *et al*., 2021], S11 [Kim *et al*., 2023], S12 [Saoud and Romman, 2024], S13 [Putra Antonio *et al*., 2023], S14 [Balaji and Rao, 2024], S15 [Rani *et al*., 2024], S16 [Ding and Najaf, 2024], S17 [Verma *et al*., 2024], S18 [Liu *et al*., 2022], S31 [Haugeland *et al*., 2022], S41 [Sauka *et al*., 2026], S54 [Hasselwander and Lah, 2026]
Education	Studies that analyse chatbots in educational settings, including learning outcomes, student experiences, academic integrity, classroom support, and AI-assisted learning processes.	S19 [Rodríguez-Ruiz *et al*., 2025], S20 [Virvou and Tsihrintzis, 2024], S21 [Jain *et al*., 2024], S22 [Kothari and Noel Newton Fernando, 2024], S23 [Karataş *et al*., 2024], S24 [Wicaksono *et al*., 2024], S35 [Wang *et al*., 2022], S37 [Ludin *et al*., 2022], S40 [Adnin *et al*., 2025], S44 [Zhao, 2025], S47 [Marmolejo-Ramos *et al*., 2026], S48 [Atcheson *et al*., 2025], S51 [Alshammari *et al*., 2025], S52 [Alshammari and Babu, 2025]
Mental Health	Studies that evaluate chatbots in mental health, well-being, or emotionally sensitive contexts, including counselling, companionship, support tools, and health-related psychological outcomes.	S25 [Yeh *et al*., 2025], S26 [Peng *et al*., 2024], S27 [Park *et al*., 2024], S28 [Siddals *et al*., 2024], S29 [Pozzi and De Proost, 2024], S30 [Maples *et al*., 2024], S32 [Kang and Hong, 2025], S33 [Shan *et al*., 2022], S34 [He *et al*., 2022], S36 [Chung *et al*., 2021], S37 [Ludin *et al*., 2022], S38 [Wong *et al*., 2021], S39 [Chagas *et al*., 2023], S53 [Jung *et al*., 2025]

Table 3. Context x Chatbot type

Context x Chatbot type	ChatGPT	Amazon's AI	General	Others
General AI-HCI	S1 [Fabio *et al*., 2024]		S2 [Akhtar *et al*., 2024], S4 [Joshi and Bengler, 2024], S8 [Weber and Ludwig, 2020]	
Design of chatbots interfaces	S32 [Kang and Hong, 2025]		S5 [Meng *et al*., 2024], S10 [Aumüller *et al*., 2024], S15 [Rani *et al*., 2024], S31 [Haugeland *et al*., 2022]	S3 (ChatrEx [Khurana *et al*., 2021], S6 (EmoBot) [Mubassira *et al*., 2025]
Emotional				S6 (EmoBot) [Mubassira *et al*., 2025]
Relationship			S7 [Zimmerman *et al*., 2024]	
E-commerce/ Consumer		S11 [Kim *et al*., 2023], S14 [Balaji and Rao, 2024]	S9 [Min *et al*., 2021], S12 [Saoud and Romman, 2024], S13 [Putra Antonio *et al*., 2023], S15 [Rani *et al*., 2024], S16 [Ding and Najaf, 2024], S18 [Liu *et al*., 2022], S41 [Sauka *et al*., 2026], S54 [Hasselwander and Lah, 2026]	S11 (Imaginary AI Agent) [Kim *et al*., 2023], S17 (Voice-activated chatbots) [Verma *et al*., 2024]
General education use	S19 [Rodríguez-Ruiz *et al*., 2025], S22 [Kothari and Noel Newton Fernando, 2024], S23 [Karataş *et al*., 2024], S40 [Adnin *et al*., 2025], S51 [Alshammari *et al*., 2025], S52 [Alshammari and Babu, 2025], S48 [Atcheson *et al*., 2025], S47 [Marmolejo-Ramos *et al*., 2026], S44 [Zhao, 2025]		S19 [Rodríguez-Ruiz *et al*., 2025], S20 [Virvou and Tsihrintzis, 2024], S21 [Jain *et al*., 2024], S24 [Wicaksono *et al*., 2024]	
Trust			S9 [Min *et al*., 2021], S16 [Ding and Najaf, 2024], S24 [Wicaksono *et al*., 2024], S33 [Shan *et al*., 2022], S39 [Chagas *et al*., 2023]	S27 (Dialogflow) [Park *et al*., 2024], S35 (SnehAI) [Wang *et al*., 2022]
Mental health	S28 [Siddals *et al*., 2024], S30 [Maples *et al*., 2024], S53 [Jung *et al*., 2025]		S26 [Peng *et al*., 2024], S33 [Shan *et al*., 2022]	S25 (Woebot app) [Yeh *et al*., 2025], S27 (Dialogflow) [Park *et al*., 2024], S28 (Pi and Kindroid) [Siddals *et al*., 2024], S29 (Karim) [Pozzi and De Proost, 2024], S34 (XiaoE) [He *et al*., 2022], S36 (Dr. Joy) [Chung *et al*., 2021], S37 (Aroha) [Ludin *et al*., 2022], S38 (ClaimIt) [Wong *et al*., 2021]
Behavior	S45 [Wang *et al*., 2025], S50 [Mayerhofer *et al*., 2025]		S42 [Genç *et al*., 2025], S43 [Shahid *et al*., 2025], S46 [He *et al*., 2022], S49 [Kang *et al*., 2025], S55 [Tsumura and Yamada, 2025], S56 [Thalhammer, 2025]	

domain-specific anthropomorphic designs. Within the educational context, we found fourteen articles [S19, S20, S21, S22, S23, S24, S35, S37, S40, S44, S47, S48, S51, S52]. Eight of them [S19, S22, S23, S40, S44, S48, S51, S52] specifically focused on ChatGPT, while all the others addressed the use of chatbots in general, with the exceptions of articles S35 and S37, which deal with a specific type of chatbot. Meanwhile, seven studies [S3, S5, S6, S10, S15, S31, S32] focused on the design of chatbot interfaces. Four of them addressed chatbots in general [S5, S10, S15, S31]; one used ChatrEx [S3], a specific type of chatbot designed to help evaluate the design of this type of tool; another article [S6] used EmoBot, an emotionally aware chatbot designed to detect and respond to users' emotions; and another [S32] used ChatGPT 4.0. The authors [S6] sought to understand how people interacted with the platform and looked for improvements for a better UX. Beyond OpenAI, five studies [S40, S45, S49, S54, S56] have begun to explore the impact of alternative models, including Google Gemini, Microsoft Copilot, Claude, and custom-built agents like ExplainitAI.

Six studies explored user trust in chatbots [S9, S16, S24, S27, S33, S35, S39], and five of them [S9, S16, S24, S33, S39] did not specify which type of chatbot was analysed. Two studies [S9, S16] about trust addressed the topic within the context of consumption, both of which seek to understand the best ways to make consumers feel comfortable and provide good service. Another study [S24] examined it within the educational context, aiming to discover whether higher education students are able to use the tool for college assignments without compromising their academic integrity. The authors also seek to show how students can trust chatbots. One study [S27] explored feelings of trust in the context of mental health in order to see how confident people feel about receiving medical recommendations from a chatbot compared to when they receive them from another human.

Two articles [S11, S14] used Amazon's AI as a reference, both within a commercial context. One of them [S11] also used an imaginary AI agent for the study. Another article on the same topic used voice-activated chatbots; however, most studies in this category [S9, S12, S13, S15, S16, S18, S54] have not delved into any specific type of chatbot.

One study in particular [S7] focused on the relationship between humans and AI. The authors addressed a wide range of forms and impacts that this relationship can have. No specific type of chatbot was the focus of the study.

Focusing on articles related to mental health, we found that nine out of twelve studies discussed specific types of chatbots [S25, S27, S28, S29, S30, S34, S36, S37, S38, S53]. Three of them [S28, S30, S53] delved into OpenAI's ChatGPT, as can be seen in Table 3. Two studies [S26, S33] do not focus on any type of chatbot, approaching them in a general way and seeking to understand their impact on social interaction, especially among lonely individuals.

Eight studies [S42, S43, S45, S46, S49, S50, S55, S56] investigate how users perceive and interact with AI, exploring conversational explainable AI (XAI) interfaces [S46], cultural perceptions of explainability [S49], and multimodal accessibility tools [S56]. One study [S45] evaluates and compares user performance in ChatGPT and Google Gemini, while another study [S55] focuses on computer-animated agents designed for reflective support, trust, and empathy.

5.2 RQ2: What research objectives are addressed in the analysed studies?

For this RQ, we analysed the purpose of each of the 56 reviewed studies. These studies address a wide array of topics, resulting in seven distinct categories: Ethical Aspects, Chatbots in Education, Chatbots in Healthcare & Well-Being, Chatbot Avatars & User Perception, Chatbots & Customer Experience, Enhancing AI User Interaction, and Explainable Chatbots.

The Ethical Aspects category consists of studies that mention the ethical aspects of human-chatbot interaction and is divided into three subcategories: Conversational AI: UX & Ethics, Chatbot Trust & Adoption Drivers, and Chatbots & Epistemic Injustice. Conversational AI: UX & Ethics includes four studies [S2, S4, S6, S8] that explore how NLP and ethical considerations influence user experience, propose adaptive human-AI interaction models, and set design quality criteria. Chatbot Trust & Adoption Drivers contains four studies [S9, S12, S24, S27], examining factors such as attention to the needs of users, empathy, contextual comprehension, and perceived ease of use that build user trust and drive chatbot adoption, including the development of trust frameworks for various sectors. One study [S27] examined a serial mediation mechanism to test whether the human representation of chatbots affects people's trust in them and whether they follow their recommendations more closely. Finally, Chatbots & Epistemic Injustice is represented by three studies [S29, S40, S56] that address the harm to individual humans as knowers and to users' ability to understand, framing it as a form of participatory injustice caused by chatbots [S29], also mentioning the 'shadow' side of AI and academic integrity regarding GenAI [S40], as well as accessibility and inclusive design for users with disabilities [S56].

Chatbots in Education include eleven studies [S1, S19, S21, S22, S23, S25, S30, S44, S47, S51, S52] exploring the role of chatbots in human-computer interaction within educational settings, examining their theoretical foundations, applications, and challenges, with a focus on language learning, critical thinking, and group counselling. A study [S19] analysed how factors such as self-control, self-esteem, and self-efficacy influence the use of AI tools among university students, while others [S22, S23, S1] investigated ChatGPT's impact on decision-making, privacy concerns, and learning outcomes. Together, these studies offer information on the benefits, such as facilitating the creation of learning materials, engaging students through interactive questioning, and the overall enrichment of the teaching process; as well as limitations, such as identifying deep structural and pragmatic errors in learners' writing and enhancing language skills. One study [S47] examined cross-national adoption patterns, another study [S44] focused on specific professional contexts like medical education, and two other studies [S51, S52] conducted structural analyses using acceptance models.

Chatbots in Healthcare & Well-Being include ten studies [S32, S33, S34, S35, S36, S37, S38, S39, S50, S53] that focus on chatbots in health-related contexts. Two studies [S33, S39] evaluated the quality of user interaction with COVID-19 response chatbots, examining overall UX, strengths, and out-

comes in both real-world and simulated scenarios. Similarly, other studies [S32, S36, S38] conducted usability and feasibility assessments of chatbot prototypes designed for clinical, perinatal, or sensitive family contexts, aiming to inform clinical studies by identifying gaps and behavioural intentions. Two studies [S50, S53] compared the use of LLMs in daily mental health support and health-related information-seeking with traditional search interfaces. One study [S36] tested both the clinical effectiveness and nonclinical performance of a CBT-based mental health chatbot, while another Study [S35] applied Gibson's theory of affordances to examine how a chatbot can educate adolescents and young adults on sexual and reproductive health.

Chatbot Avatars & User Perception has nine studies [S7, S10, S25, S26, S28, S42, S43, S49, S55], which aim to explore how users perceive and interact with chatbot avatars, particularly focusing on gender representation, personalization, and psychological factors. They examine how design elements, such as anthropomorphic or gender-neutral features, influence acceptance, customisation choices, and continued use. In addition, they investigate how user characteristics, such as gender and loneliness, affect preferences and usage intentions for conversational AI, especially in emotional and mental health contexts.

Chatbots & Customer Experience comprises seven studies [S11, S13, S14, S15, S16, S18, S31] that examined how chatbot interactions influence customer satisfaction, loyalty, and brand advocacy. They highlighted the roles of usability, personalization, interface design, and trust, particularly in service bots and robo-advisors. Additionally, these studies investigated how non-human agents in service settings affect unethical behaviour and how interactivity and perceived humanness influence trust and adoption intentions in e-commerce, with trust acting as a mediating factor in these relationships.

Enhancing AI User Interaction has four studies [S5, S17, S45, S46] focused on enhancing AI-assisted user interactions. One developed design features to support flexible, goal-orientated query formulation without relying on rigid prompts, while the other examined the benefits and challenges of voice-activated chatbots across industries and their impact on customer experience.

Lastly, Explainable Chatbots present a study [S3] that explores the design of explainable chatbot interfaces in an application that explains the underlying workings of a chatbot during a breakdown.

In Table 4, we can visualise all categories along with each of their subcategories, objectives, and articles; in Figure 5, we can see the number of articles in each category for RQ2.

5.3 RQ3: How are human–chatbot interactions characterised and evaluated in the literature, and what user perceptions and experiences are reported across different application contexts?

In this RQ, we explored the way users perceive and interact with AI systems according to the 56 reviewed studies. These studies cover an extensive range of topics that we grouped into seven diverse categories, each related to the context of use of the mentioned chatbot, as the manner in which the interaction was carried out differed in each context of use: Ethical Aspects, Chatbots in Education, Chatbots & Customer Experience, Chatbot Avatars & User Perception, Enhancing AI User Interaction, AI for Data Query & Insights, and Participatory Injustice in AI. These categories result from subdividing the research based on the specific context of human-AI interaction observed within the studies, derived inductively through iterative coding based on the context of use in which each study evaluated its chatbot. These initial codes were then grouped into higher-level categories by comparing similarities across studies and merging similar contexts.This approach allows us to delve deeper into how the interplay between users and AI systems manifests in different application areas.

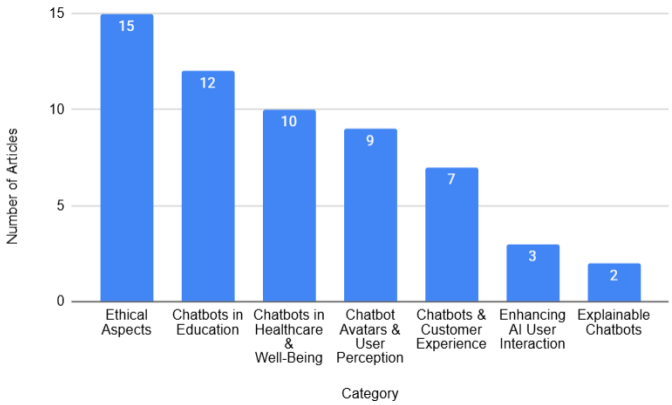

Figure 5. Number of articles per category

The Ethical Aspects category is divided into five subcategories: AI in Health, Conversational AI: UX & Ethics, Chatbot Trust & User Acceptance, Chatbot Trust & Adoption Drivers, and AI Ethics: Education & Professional Features. AI in Health encompasses sixteen studies [S4, S8, S11, S25, S27, S30, S32, S33, S34, S35, S36, S37, S38, S39, S50, S53] that focus on advancing the design and understanding of conversational AI, with a notable emphasis on its applications and ethical implications within the health domain. They investigate how to optimise UX through adaptive interaction strategies, which are particularly relevant for digital health assistants used in symptom checking or preliminary diagnosis. They also describe how interaction was carried out and how it was experienced by users in real-world and experimental healthcare settings. The main aspects of natural communication explored are sensible questioning behaviour for breakdowns, contextual understanding, the ability to handle complex queries, and natural voice output, which are important factors for effective, empathetic, and healthy interactions. Their research also explores the ethical implications of AI interaction, including how human-like AI can influence user behaviour and trust, even in sensitive health contexts. Studies examine the psychological impact of perceived human representation in chatbots, investigating how it affects trust and compliance with health recommendations, particularly from chatbot counsellors. Furthermore, research explores the role of AI tools in reducing anxiety and depression and nuanced user relationships with AI companions, highlighting their perceived social support for lonely or vulnerable individuals. Conversational AI: UX & Ethics encompasses six studies [S2, S3, S17, S40, S41, S46], focusing on advancing conversational AI through qualitative methodologies

Table 4. RQ2 - Categories, Subcategories, Objectives, and Articles

Category	Subcategories	Objective	Articles
Ethical Aspects	Conversational AI: UX & Ethics	Explore how NLP and Ethics influence UX, proposing adaptive human-AI interaction models and setting design quality criteria.	S2 [Akhtar *et al*., 2024], S4 [Joshi and Bengler, 2024], S6 [Mubassira *et al*., 2025], S8 [Weber and Ludwig, 2020], S40 [Adnin *et al*., 2025]
	Chatbot Trust & Adoption Drivers	Explore drivers of trust and adoption, such as empathy, user needs, ease of use, contextual understanding and Human Representation of Chatbots	S9 [Min *et al*., 2021], S12 [Saoud and Romman, 2024], S24 [Wicaksono *et al*., 2024], S27 [Park *et al*., 2024], S41 [Sauka *et al*., 2026], S54 [Hasselwander and Lah, 2026]
	Chatbots & Epistemic Injustice	Address the epistemic harm caused by chatbots, affecting users' ability to understand and act as knowers.	S29 [Pozzi and De Proost, 2024], S56 [Thalhammer, 2025]
Chatbots in Education	–	Explore the role of chatbots in human-computer interaction within educational settings, examining their theoretical foundations, applications, and challenges, with a focus on language learning, critical thinking, and group counselling	S1 [Fabio *et al*., 2024], S19 [Rodríguez-Ruiz *et al*., 2025], S21 [Jain *et al*., 2024], S22 [Kothari and Noel Newton Fernando, 2024], S23 [Karataş *et al*., 2024], S25 [Yeh *et al*., 2025], S30 [Maples *et al*., 2024], S44 [Zhao, 2025], S47 [Marmolejo-Ramos *et al*., 2026], S48 [Atcheson *et al*., 2025], S51 [Alshammari *et al*., 2025], S52 [Alshammari and Babu, 2025]
Chatbots in Healthcare & Well-Being	–	Investigate chatbots in health-related contexts, including COVID-19 response chatbots and prototypes designed for clinical, perinatal, or sensitive family settings. Assess usability, feasibility, and behavioural intentions to inform future clinical studies, and evaluate both clinical effectiveness and nonclinical performance in mental health and educational health applications.	S32 [Kang and Hong, 2025], S33 [Shan *et al*., 2022], S34 [He *et al*., 2022], S35 [Wang *et al*., 2022], S36 [Chung *et al*., 2021], S37 [Ludin *et al*., 2022], S38 [Wong *et al*., 2021], S39 [Chagas *et al*., 2023], S50 [Mayerhofer *et al*., 2025], S53 [Jung *et al*., 2025]
Chatbot Avatars & User Perception	–	Study how users perceive chatbot avatars, including aspects such as gender representation, anthropomorphism, and psychological impact.	S7 [Zimmerman *et al*., 2024], S10 [Aumüller *et al*., 2024], S25 [Yeh *et al*., 2025], S26 [Peng *et al*., 2024], S28 [Siddals *et al*., 2024], S42 [Genç *et al*., 2025], S43 [Shahid *et al*., 2025], S49 [Kang *et al*., 2025], S55 [Tsumura and Yamada, 2025]
Chatbots & Customer Experience	–	Examine how chatbot interactions influence customer satisfaction, loyalty, and brand advocacy. Emphasize the roles of usability, personalization, interface design, interactivity, perceived humanness, and trust (including trust as a mediating factor), while also investigating how non-human agents in service settings may affect unethical behaviour and adoption intentions.	S11 [Kim *et al*., 2023], S13 [Putra Antonio *et al*., 2023], S14 [Balaji and Rao, 2024], S15 [Rani *et al*., 2024], S16 [Ding and Najaf, 2024], S18 [Liu *et al*., 2022], S31 [Haugeland *et al*., 2022]
Enhancing AI User Interaction	–	Develop and evaluate methods to improve user interaction with AI, such as voice interfaces, without relying on rigid prompts	S5 [Meng *et al*., 2024], S17 [Verma *et al*., 2024], S45 [Wang *et al*., 2025]
Explainable Chatbots	–	Explore the design of explainable chatbot interfaces in an application that explains the underlying workings of a chatbot during a breakdown.	S3 [Khurana *et al*., 2021], S46 [He *et al*., 2025]

(document reviews, interviews, open-ended questionnaires) and technical advancements (keyword extraction, contextual understanding) to enhance language comprehension and response relevance, alongside usability studies on transparency and trust for explainable AI interfaces. They noticed that chatbots have demonstrated their value and effectiveness in tasks including customised recommendations, task automation, and information retrieval. However, at the same time, users were more likely to lose trust and less likely to continue using these chatbots after experiencing a breakdown, especially when engaged in complex tasks. Chatbot Trust & User Acceptance includes two studies [S16, S28], exploring how high interactivity and satisfied user expectations foster trust and adoption, noting positive user perceptions (e.g., viewing chatbots as an 'emotional sanctuary') despite concerns like privacy and technological immaturity. Chatbot Trust & Adoption Drivers contains one study [S24], which quantitatively investigates factors influencing student trust in chatbots using Structural Equation Modeling (SEM) and questionnaire-based data collection. Lastly, AI Ethics: Education & Professional features one study [S20] that introduces a comprehensive framework designed to optimise AI performance within educational contexts, with a primary focus on ethical responsibility, protecting privacy, and ensuring secure data collection, storage, and usage through robust data governance policies. As in education, social robots with affective capabilities can detect students' emotional states and adapt their teaching methods to better suit individual needs. This ability to respond to emotional feedback can create more engaging and supportive learning environments, helping students feel understood and motivated. Robots that respond to emotions within specific contexts can enhance human-robot interactions.

Chatbots in Education comprises seven studies [S19, S23, S44, S47, S48, S51, S52] that investigate the impact and implications of chatbots, particularly ChatGPT, within educational settings. These studies aim to explore how AI affects students' learning experiences, including their engagement, motivation, and skill acquisition in various subjects, notably foreign language learning. Research in this area examines both the strengths and challenges of integrating AI into education, collecting student perspectives on its perceived advantages and limitations, as well as suggestions for effective implementation. Methodologies often involve quantitative approaches, such as cross-sectional ex-post facto designs using self-reported questionnaires among university students. Some studies specifically assess the influence of AI familiarity on critical thinking abilities, highlighting the potential for conscientious AI interaction to improve cognitive skills. Additionally, this category addresses ethical concerns related to AI use in education, such as student vulnerability due to a lack of training and increasing concerns about safety and privacy with widespread adoption.

Chatbots & Customer Experience covers five studies [S13, S14, S18, S31, S54] that quantitatively assess how chatbot interactions influence customer satisfaction and behaviour. These studies used online surveys to collect data from online shoppers, focusing on aspects such as ease of use, response speed, and information accuracy. They focused on identifying significant relationships between chatbot features and user satisfaction using regression analysis. Although the research backgrounds vary and extensively analyse UX, there are no explicit ethical considerations or detailed user perceptions beyond satisfaction.

Chatbot Avatars & User Perception includes five studies [S10, S26, S42, S43, S55] exploring how user perception and interaction with chatbot avatars influence acceptance and usage, with a particular focus on gender representation, anthropomorphism, and the emotional needs of users. Studies indicate that the perceived gender of an avatar can significantly impact the user's intention to use a chatbot. Additionally, the research questions suggest that although anthropomorphic or gender-neutral designs might aim to prevent stereotypical representations, empirical evidence indicates that human-like qualities may lead to higher user acceptance and foster emotional bonds, especially for lonely individuals seeking social connections.

Enhancing AI User Interaction includes three studies [S6, S15, S45] that synthesise existing knowledge and propose design adjustments. While one study [S15] reviews interactions from various systems to provide a comprehensive overview of HCI concepts, drawing from textbooks and expert-created materials, the other [S6] directly enhances UX by assessing pre-interaction expectations and post-interaction feedback with emotion-generating chatbots. The goal is to identify the gaps between user desires and actual AI performance, suggesting specific adjustments to improve alignment and the overall HCI experience.

AI for Data Query & Insights includes a study [S5] focused on improving AI-assisted user interactions, specifically within the domain of log management and security monitoring. It emphasises the derivation of information and functional requirements through interviews with domain experts, highlighting the need to enable users to efficiently search, filter, and analyse log data by using relevant keywords, platform-native syntax, timeframes, and search categories, ultimately aiming to improve the way users query and gain insight from complex system data.

Furthermore, Participatory Injustice in AI includes a study [S29] that critically examines epistemic injustice within AI development, specifically focusing on participatory injustice. The study highlights a case where individuals were treated merely as "testing subjects" rather than as potential participants in discussions about application development.These findings, based on prejudice and stereotypes, can lead to individuals being epistemically disempowered or "disabled" when their contributions are not taken seriously, especially regarding mental health concerns.

In Table 5, we can visualise all categories along with each of their subcategories, objectives, and articles. In Figure 6, we can see the number of articles in each category for RQ3.

Additionally, we will look at how UX is rated across the systems mentioned in each study. These studies cover a range of topics, leading to eight distinct categories: Mixed User Evaluation and Framework Assessment, Rated UX and Demographic Analysis, Experimental Designs and Diverse Samples, Outcome-Based Performance Assessment, Comprehensive Demographic and Performance Assessment, Quantitative Assessment, AI Impact on User Well-being and Behaviour, and Usability Evaluation for Different User Groups. It is important to highlight that some studies did not provide any evidence

Table 5. RQ3 – Thematic Categorization of Human-AI Interaction Contexts and Study Objectives

Category	Subcategories	Objective	Articles
Ethical Aspects	AI in Health	Investigate the design, ethics, and interaction strategies of conversational AI in health contexts, focusing on empathy, breakdowns, trust, mental health impacts, and real-world user experience in clinical and public health settings.	S4 [Joshi and Bengler, 2024], S8 [Weber and Ludwig, 2020], S11 [Kim *et al*., 2023], S25 [Yeh *et al*., 2025], S27 [Park *et al*., 2024], S30 [Maples *et al*., 2024], S32 [Kang and Hong, 2025], S33 [Shan *et al*., 2022], S34 [He *et al*., 2022], S35 [Wang *et al*., 2022], S36 [Chung *et al*., 2021], S37 [Ludin *et al*., 2022], S38 [Wong *et al*., 2021], S39 [Chagas *et al*., 2023], S50 [Mayerhofer *et al*., 2025], S53 [Jung *et al*., 2025]
	Conversational AI: UX & Ethics	Advance conversational AI through qualitative and technical methods to improve comprehension and trust, highlighting breakdown effects on user perception.	S2 [Akhtar *et al*., 2024], S3 [Khurana *et al*., 2021], S17 [Verma *et al*., 2024], S40 [Adnin *et al*., 2025], S41 [Sauka *et al*., 2026], S46 [He *et al*., 2025]
	Chatbot Trust & User Acceptance	Explore how interactivity and fulfilled expectations build trust and adoption, even in emotionally sensitive uses.	S16 [Ding and Najaf, 2024], S28 [Siddals *et al*., 2024], S41 [Sauka *et al*., 2026]
	Chatbot Trust & Adoption Drivers	Quantitatively investigate factors influencing student trust in chatbots using Structural Equation Modeling (SEM) and questionnaire-based data collection.	S24 [Wicaksono *et al*., 2024]
	AI Ethics: Education & Professional	Introduce an ethical framework for educational AI, emphasising privacy, governance policies, and emotional adaptability.	S20 [Virvou and Tsihrintzis, 2024]
Chatbots in Education	–	Examine AI's influence on student engagement, motivation, learning outcomes, and ethical concerns in educational settings.	S19 [Rodríguez-Ruiz *et al*., 2025], S23 [Karataş *et al*., 2024], S44 [Zhao, 2025], S47 [Marmolejo-Ramos *et al*., 2026], S48 [Atcheson *et al*., 2025], S51 [Alshammari *et al*., 2025], S52 [Alshammari and Babu, 2025]
Chatbots & Customer Experience	–	Analyze how chatbot interactions influence customer satisfaction and behaviour, using quantitative and experimental methods.	S13 [Putra Antonio *et al*., 2023], S14 [Balaji and Rao, 2024], S18 [Liu *et al*., 2022], S31 [Haugeland *et al*., 2022], S54 [Hasselwander and Lah, 2026]
Chatbot Avatars & User Perception	–	Explore how chatbot avatar features, such as gender and anthropomorphism, affect acceptance and emotional connection, especially for socially isolated users.	S10 [Aumüller *et al*., 2024], S26 [Peng *et al*., 2024], S42 [Genç *et al*., 2025], S43 [Shahid *et al*., 2025], S55 [Tsumura and Yamada, 2025]
Enhancing AI User Interaction	–	Identify UX gaps by synthesising interaction studies and emotional feedback to guide design improvements in HCI.	S6 [Mubassira *et al*., 2025], S15 [Rani *et al*., 2024]
AI for Data Query & Insights	–	Improve data query efficiency through expert-informed design for log management and security monitoring systems.	S5 [Meng *et al*., 2024]
Participatory Injustice in AI	–	Examine how epistemic injustice occurs in AI development when individuals are treated merely as "testing subjects", especially in mental health-related contexts.	S29 [Pozzi and De Proost, 2024]

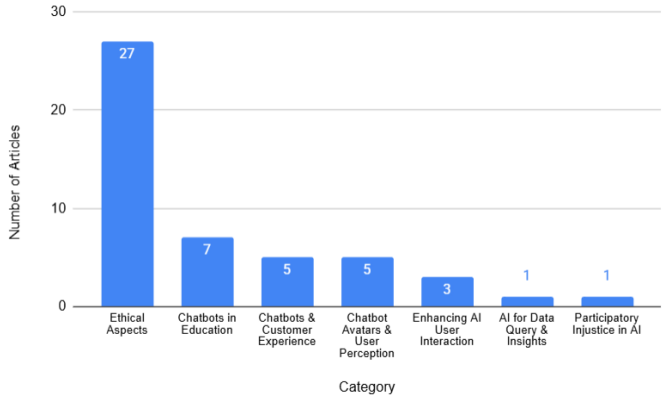

Figure 6. number of articles per category

that the evaluation process was carried out, and many did not mention the number of participants involved; it was only stated that the evaluation process was conducted, but no information was given about the kind or number of participants; therefore, we cannot generalise the results since they would not be valid for use in other studies.

Starting with Mixed User Evaluation and Framework Assessment, it encompasses eleven studies [S2, S3, S9, S11, S12, S20, S23, S32, S39, S45, S48] that use various methods to assess UX, often revealing mixed evaluations of AI tools. One study [S9] found ambivalent consumer opinions on AI acceptance; another study [S20] assessed AI performance against pedagogical principles through student feedback. Five studies [S2, S3, S11, S12, S23] used various data collection and analysis techniques: one study [S12] employed a five-point Likert scale in an online survey, and another [S2] described a comprehensive research methodology integrating ethical and UX considerations. Similarly, one study [S39] employed a convergent parallel mixed methods design, using both a post-task usability survey with real chatbot users and interviews based on simulated interactions to explore UX. Another study [S32] also used a mixed methods design by combining structured quantitative items with qualitative questions on strengths and areas for improvement.Three other studies [S3, S11, S23] combined statistical tests with qualitative analysis to explore user perceptions of usefulness, transparency, and trust, particularly regarding breakdown recovery and critical thinking. These studies often involved diverse student populations, analysing demographic data to understand varied experiences. Two studies S45, S48] employed multi-phase interviews and long-term development meetings to capture user perceptions of GenAI ecosystems and disability-related impacts.

In the Rated UX and Demographic Analysis, nine studies [S6, S10, S19, S26, S36, S38, S41, S46, S55] quantify UX through explicit rating systems and provide detailed demographic breakdowns of participants. These studies used different methods, such as a 0-10 scale for chatbot performance [S6], five-point Likert scales for aspects such as conversational AI usage and anthropomorphism [S26], and self-reported questionnaires to assess user traits and AI tool use [S19]. They also gathered extensive demographic data, including age, gender, education, and occupation of various populations in different countries. In healthcare-related settings, one study [S36] evaluated usability using the USE (usefulness, ease of use, ease of learning, and satisfaction) questionnaire, reporting that ease of learning achieved the highest mean score, followed by ease of use, satisfaction, and usefulness. Similarly, another study [S38] assessed UX through a chatbot usability questionnaire, where parents rated ease of use and overall satisfaction on a five-point scale, supported by qualitative feedback describing the chatbot as a successful and positive tool. One study [S55] uses validated empathy and trust questionnaires to measure behavioral adaptation, while two other studies [S41, S46] use experimental designs to quantify how explainability and anthropomorphism correlate with trust in high-stakes environments.

In Experimental Designs and Diverse Samples, including seven studies [S16, S25, S29, S30, S37, S42, S43,],various experimental and sampling designs are used to evaluate UX and behaviour, often involving diverse populations and contexts. Studies evaluated the impact of chatbots on group counselling [S25] using a quasi-experimental design with master's students and analysed interaction patterns across various demographics in general user bases [S16]. It also includes research with vulnerable populations [S29] using a chatbot pilot and university students using AI companions such as Replika [S30], providing information on various UXs in different real-world scenarios. Similarly, one study [S37] evaluated a chatbot in a real-world rollout during the COVID-19 pandemic, using online questionnaires and a focus group with community members. The real-world setting also enabled iterative engineering based on actual user inputs, demonstrating how evaluation and design refinement can occur simultaneously in deployed systems. One study [S42] conducted crowd-sourced studies and another [S43], longitudinal in-home deployments with older adults.

Outcome-Based Performance Assessment features four studies [S17, S35, S51, S52]. One study [S17] evaluates the UX through the lens of model outcomes, including accuracy, consumer preferences, and management actions based on the information submitted by the respondents. Another study [S35] evaluates a CBT-based mental health chatbot using an outcome-driven design with multiple measurement time points, also measuring secondary outcomes such as usability, acceptability, and working alliance. One study [S51], used SEM to evaluate how readiness, interactivity, and ethical awareness drive AI adoption, while another [S52] measured how satisfaction mediates the relationship between usefulness and intention to use AI.

For Comprehensive Demographic and Performance Assessment, it comprises three studies [S1, S11, S27] that feature extensive demographic data along with performance evaluations or specific behavioural outcomes related to AI interaction. These studies often involved multiple experiments or groups of different participants. For example, a study [S11] recruited undergraduate students and online workers to examine different behaviours, such as predicting numbers or product returns with AI. Another study [S27] with 385 participants evaluated responses to chatbot representations, including data on education and income. Lastly, a study [S1] with 126 university students used an objective assessment (CRA) to evaluate critical thinking abilities.

In Quantitative Assessment, two studies [S14, S18] quantitatively evaluate UX, primarily through hypothesis testing

and statistical analysis. They use methods such as descriptive statistics and multiple linear regression to understand the relationships between AI, chatbot features, and customer satisfaction. Although these studies perform rigorous statistical analyses on user data, they do not explicitly mention a specific rating system or scale used to capture user feedback.

In AI Impact on User Well-being and Behaviour, two studies [S7, S28] explore the complex impacts of chatbots on user well-being, relationships, and behaviour, moving beyond simple rating systems. One study [S7] discusses potential dangers, such as anthropomorphism, lack of personal connection, and the encouragement of unhealthy behaviours. It also examines human-AI relationships, noting concerns over negative impacts from uncontrolled AI personalities while highlighting benefits such as the alleviation of loneliness for vulnerable groups. Another study [S28] revealed that while users often found mental health chatbots to be understandable, AI's "safety guardrails" could disrupt their emotional sanctuary, leading to feelings of unpleasantness or rejection, emphasising how design choices in sensitive areas directly affect the users' emotional experience.

Lastly, in the Usability Evaluation for Different User Groups, the study [S5] performed a usability evaluation to assess the effectiveness and satisfaction of design features. This evaluation explicitly considered differing findings for experienced and novice users, reflecting technical diversity. The evaluators included data analysts with prior domain experience in log management and a novice graduate student with usability experience, allowing for a comprehensive assessment across different levels of user familiarity. Another study [S56], proposed a Research-through-Design (RTD) methodology for accessibility-focused multimodal interfaces.

In Table 6, we can see all the studies that have specified the number of participants in their evaluation process, together with the way in which UX was rated, the type of study, and each article ID; in Table 7, we can visualise all categories with their objectives and articles, and in Figure 7, we can see the number of articles in each category for RQ3.

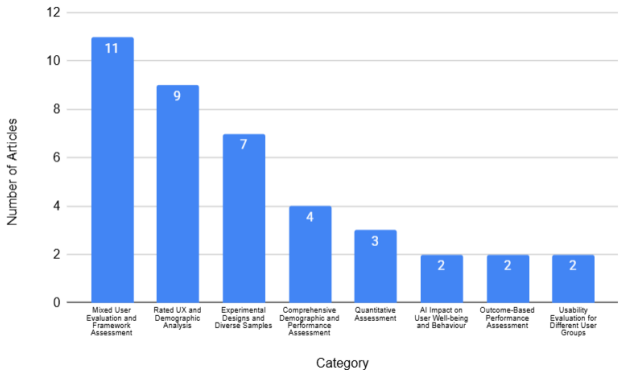

Figure 7. Number of articles per category

5.4 RQ4: What impacts of chatbot use on users' behaviour and everyday life are reported across studies?

This RQ aims to understand how chatbots influence users' lives through the analysis and conclusions obtained from the 56 reviewed articles. The categories in this RQ were derived inductively through iterative coding, based on the context of use in which each study evaluated its chatbot. The studies were first coded at the descriptive level by identifying the chatbot's application context and then grouped into categories by comparing similarities across studies.

One study [S1] discussed how the use of ChatGPT can be directly associated with critical thinking skills. The study comprised only university students, but its focus is not limited to the educational environment; it also expands to the impact of the platform in the work environment, for example. The authors concluded that using the platform helps user's critical thinking and is further enhanced when used with a critical mindset and caution. In addition, this engagement is associated with a more in-depth analysis of the feedback from ChatGPT responses, reducing user's dependence on the platform. The study also highlighted the importance of ChatGPT in helping students engage in their studies due to its personalised and immediate responses, as well as in engaging workers in the workplace by assisting them with relevant solutions and data. The use of chatbots can also help companies not only by assisting their employees but also by dealing directly with their customers. A reviewed study [S9] concluded that once customers are familiar with using these tools, they can be applied, with the use of AI, directly to customer service to provide better service.

It is also important to analyse how users interpret chatbots. According to a study [S10], the use of anthropomorphic characteristics in CAs is perceived more positively by people compared to chatbots with characteristics attributed to a specific gender, especially male ones, which are perceived less favourably. The way users perceive and engage with these tools also impacts their behaviour, according to another study [S11]. The authors state that there is a greater tendency for people to behave unethically towards a human due to a lack of guilt, and this can be mitigated by developing the individual's relationship with the AI.

Interface design also shapes the hedonic and pragmatic experience, influencing user satisfaction. An experimental study [S39] demonstrated that using buttons instead of free text input can reduce cognitive load, increasing the feeling of control and efficiency, which positively impacts the perceived usefulness of the tool in everyday life.

Two studies [S14, S15] discuss ways to improve chatbots to ensure a good UX. According to one of these studies [S14], it is necessary to improve the NLP of these platforms so that users continue to use them and have higher levels of satisfaction. The other study [S15] addresses the importance of good interface design and usability, which are critical factors for long-term engagement.

In addition to the previously mentioned study [S11], four other studies [S16, S27, S54, S55] raised issues regarding user trust. All five studies revealed that the greater the user's engagement with the chatbot, the greater their trust in the tool. However, while one study [S11] addressed the user's bad intentions towards the platform, the others [S16, S27] examined this relationship in a more positive way, promoting a greater intention for adoption and acceptance. Two recent studies extend this discussion: [S55] showed that an agent's positive behaviour modification and consistently positive attitude significantly increased both empathy and trust over time,

Table 6. Studies, Type of study, How UX was rated, and the number of participants in each study

Article	Type	Rating	Number of participants
S1 [Fabio *et al*., 2024]	Journal	Evaluating critical thinking abilities through objective, task-based assessments, focusing on actual participant performance rather than self-reported information.	126
S10 [Aumüller *et al*., 2024]	Journal	A cross-sectional online experiment with a real chatbot interaction was conducted. Perceived avatar gender was assessed by asking participants from the anthropomorphic avatar condition which gender they would assign to the avatar.	1064
S11 [Kim *et al*., 2023]	Journal	Multiple experimental studies involving interactions with AI vs human agents to assess dishonest behaviour, moral decisions, and framing effects. Scenarios included predicting numbers for rewards, selecting dishonest coupon claims, and returning products under false pretenses.	526
S12 [Saoud and Romman, 2024]	Conference	Using the snowball technique for online surveys, participants self-administered the online questionnaire after receiving an invitation via email. The questionnaire had three sections: the first gathered demographic information, the second utilized a five-point Likert scale to assess participants' perceptions, attitudes towards chatbots, and overall user experience.	150
S19 [Rodríguez-Ruiz *et al*., 2025]	Journal	Participants were asked about different AI tool use cases (e.g., solving doubts, faking media), self-efficacy, self-control, and related factors. Analyses used FACTOR for reliability and Pearson correlations.	1761
S23 [Karataş *et al*., 2024]	Journal	Students were introduced to ChatGPT through learning experiences over a span of four weeks by the researcher as a language teacher. The qualitative data collected from the interviews were analysed using thematic analysis.	13
S25 [Yeh *et al*., 2025]	Journal	One-group pretest posttest design from the category of quasi-experimental designs, where the entire group underwent experimental intervention, and effects were assessed using pre- and post-tests.	42
S26 [Peng *et al*., 2024]	Journal	Questionnaire on the Wenjuanxing platform.	1006
S27 [Park *et al*., 2024]	Journal	Two-way conversation with a mental health chatbot counselor.	385
S28 [Siddals *et al*., 2024]	Journal	Qualitative semistructured interviews, which were analysed thematically. Participants were required to have had at least three separate conversations with an LLM-based generative chatbot on mental health and wellbeing topics, each lasting at least 20 minutes.	19
S29 [Pozzi and De Proost, 2024]	Journal	Pilot chatbot test.	60
S39 [Chagas *et al*., 2023]	Journal	Mixed methods UX evaluation using (1) a post-task usability survey with real users who interacted with the chatbot for symptom checking, and (2) interviews with volunteer participants completing simulated interactions guided by an interviewer. A convergent parallel mixed methods design was used, analysing quantitative and qualitative data separately and integrating findings at the end.	63
S32 [Kang and Hong, 2025]	Journal	Mixed methods evaluation combining 8 quantitative questions (1 overall satisfaction item and 7 chatbot performance components) with 4 qualitative questions (2 positive aspects and 2 areas for improvement). This enabled triangulation between quantitative ratings and qualitative feedback.	20
S34 [He *et al*., 2022]	Journal	Outcome-based assessment with baseline (T0), post-intervention after 1 week (T1), and 30-day follow-up (T2). UX-related evaluation included secondary outcomes such as working alliance, usability, and acceptability alongside the primary clinical outcome (PHQ-9).	148
S36 [Chung *et al*., 2021]	Journal	Usability evaluation using the USE questionnaire (usefulness, ease of use, ease of learning, satisfaction). Scores were reported as mean values for each component, capturing utilitarian and hedonic value.	15
S37 [Ludin *et al*., 2022]	Journal	UX evaluation through online questionnaires and a focus group with community members in a real-world rollout. The deployment allowed iterative engineering improvements based on actual user conversational intents.	238
S38 [Wong *et al*., 2021]	Journal	UX rated using a chatbot usability questionnaire. Parents rated ease of use and overall satisfaction on a 5-point scale, supported by qualitative feedback describing the chatbot as a positive and successful tool.	26
S41 [Sauka *et al*., 2026]	Journal	Between-subjects experiment; evaluation of trust/adoption pathways.	8
S42 [Genç *et al*., 2025]	Journal	Pre/Post-interaction survey; measuring empathy/trust post-visual priming	344
S46 [He *et al*., 2025]	Journal	Empirical study (loan approval task); binary decision-making metrics	306
S51 [Alshammari *et al*., 2025]	Journal	Quantitative survey; SEM analysis of adoption determinants	211
S52 [Alshammari and Babu, 2025]	Journal	Quantitative survey; TAM/ECM construct integration	297
S53 [Jung *et al*., 2025]	Journal	Qualitative thematic analysis (social media data)	9924
S54 [Hasselwander and Lah, 2026]	Journal	Quantitative online survey; PLS-SEM and NCA analysis	243

Table 7. RQ3 - Methodological Approaches for UX Evaluation and Performance Assessment

Category	Objective	Articles
Mixed User Evaluation and Framework Assessment	Use various methods to assess UX, often revealing mixed evaluations of AI tools.	S2 [Akhtar *et al*., 2024], S3 [Khurana *et al*., 2021], S9 [Min *et al*., 2021], S11 [Kim *et al*., 2023], S12 [Saoud and Romman, 2024], S20 [Virvou and Tsihrintzis, 2024], S23 [Karataş *et al*., 2024], S32 [Kang and Hong, 2025], S39 [Chagas *et al*., 2023], S45 [Wang *et al*., 2025], S48 [Atcheson *et al*., 2025]
Rated UX and Demographic Analysis	Quantify UX using rating systems and gather detailed participant demographic data to analyse interaction patterns across populations.	S6 [Mubassira *et al*., 2025], S10 [Aumüller *et al*., 2024], S19 [Rodríguez-Ruiz *et al*., 2025], S26 [Peng *et al*., 2024], S36 [Chung *et al*., 2021], S38 [Wong *et al*., 2021], S41 [Sauka *et al*., 2026], S46 [He *et al*., 2025], S55 [Tsumura and Yamada, 2025]
Experimental Designs and Diverse Samples	Evaluate UX and behaviour through experimental and sampling designs across diverse samples including vulnerable groups, using chatbots in group counseling, education, or AI companionship.	S16 [Ding and Najaf, 2024], S25 [Yeh *et al*., 2025], S29 [Pozzi and De Proost, 2024], S30 [Maples *et al*., 2024], S37 [Ludin *et al*., 2022], S42 [Genç *et al*., 2025], S43 [Shahid *et al*., 2025]
Comprehensive Demographic and Performance Assessment	Integrate demographic data with AI performance metrics, evaluating behavioural outcomes.	S1 [Fabio *et al*., 2024], S11 [Kim *et al*., 2023], S27 [Park *et al*., 2024], S40 [Adnin *et al*., 2025], S47 [Marmolejo-Ramos *et al*., 2026]
Quantitative Assessment	Quantitatively evaluate UX, primarily through hypothesis testing and statistical analysis without using explicit rating scales.	S14 [Balaji and Rao, 2024], S18 [Liu *et al*., 2022], S51 [Alshammari *et al*., 2025], S52 [Alshammari and Babu, 2025], S54 [Hasselwander and Lah, 2026]
AI Impact on User Well-being and Behaviour	Explore the complex impacts of chatbots on user well-being, relationships, and behaviour.	S7 [Zimmerman *et al*., 2024], S28 [Siddals *et al*., 2024], S53 [Jung *et al*., 2025]
Outcome-Based Performance Assessment	Evaluate UX through model outcomes such as accuracy, preference expression, and decision-related outcomes, including usability/acceptability as secondary measures.	S17 [Verma *et al*., 2024], S34 [He *et al*., 2022], S44 [Zhao, 2025]
Usability Evaluation for Different User Groups	Perform usability tests comparing experienced and novice users, emphasising effectiveness, satisfaction, and interface familiarity.	S5 [Meng *et al*., 2024], S56 [Thalhammer, 2025]

while [S54] found that among Gen Z users, trust did not significantly predict the intention to use agentic AI — perceived usefulness was the stronger driver, even amid privacy and reliability concerns. One of these articles [S27] emphasises this behaviour in people who have more contact with chatbots and, consequently, are more likely to comply with health recommendations made by them. However, another article [S30] discusses two competing hypotheses: that chatbots replace human relationships, increasing isolation, and that they stimulate new connections; since they can fulfil the roles of a friend, a therapist, or even a mirror. This points out that chatbots have an ambiguous role in social interactions: at the same time that they mitigate loneliness, they can reinforce the distancing from concrete human relationships.

Two additional recent studies deepen this trust-adoption relationship from complementary angles. One [S47], a cross-national survey of 1906 university students across 13 countries, found that the propensity to trust AI was the strongest predictor of both positive perceptions and adoption of LLMs, outweighing the effects of AI anxiety and familiarity with algorithms. The other study [S45], approaches trust from a different view. It shows that users' mental models of how a chatbot ecosystem handles their data directly shape their trust levels: participants who understood third-party chatbot ecosystems such as ChatGPT with plugins held simpler, more consistent mental models and reported fewer privacy concerns than those reasoning about first-party ecosystems such as Google Gemini, where the blurred boundary between the chatbot and its parent company created confusion and eroded trust despite users' general confidence in large, established companies.

Adding a further layer of nuance, a study on students' adoption of AI applications in higher education [S51] found that trust and performance did not significantly predict behavioural intention, whereas readiness, interactivity, and ethical awareness did, suggesting that the weight placed on trust as a driver of adoption may vary considerably across educational contexts and user populations. Complementing this, a study using an integrated TAM–ECM model [S52] found that perceived usefulness directly and significantly predicted students' behavioural intention to use ChatGPT, whereas perceived ease of use had no direct effect; instead, satisfaction fully mediated the relationship between perceived ease of use and behavioural intention, reinforcing that students' continued use of ChatGPT depends more on how useful they find it than on how easy it is to use.

Further studies extend this discussion by examining the gap between trust perceptions and actual behaviour. Study [S42] found that an optimistic CA attitude increased users' trust, perceived competence, and closeness, but this did not translate into higher donation amounts, revealing a disconnect between trust and behavioural outcomes. Similarly, study [S46] found that conversational XAI interfaces increased users' reliance on AI-driven decisions, but this increased reliance was not necessarily appropriate; instead, it was characterised by clear patterns of over-reliance, suggesting that both interfaces functioned as persuasive technologies rather than tools that calibrate trust. Also, a study examining how users blend web search and generative AI chat for health-related information tasks [S50] found that trust was frequently misplaced in favour of ease-of-use and the seemingly polished nature of chat-generated answers, leading to increased post-search confidence even when the answer given was incorrect. Notably, sessions combining both chat and search yielded lower correctness than either tool used alone.

When addressing studies on chatbots in the educational environment [S20, S21, S23, S24], the analysis among them comes from different perspectives. Two studies [S20, S21] present an optimistic view of incorporating these systems as facilitators of student-centred learning, contributing to a more

mediating role for the teacher. The study [S23] highlighted that ChatGPT improves linguistic performance, promoting creativity and motivation. However, the authors emphasise that there is a risk of dependency, which could compromise student autonomy. The study [S24] comments that acceptance of and knowledge about the resources of chatbots increase their perceived effectiveness. Related to this behavioural dimension, study [S40] found that students frequently use GenAI without disclosing it to instructors, employing strategies such as paraphrasing AI-generated responses to avoid detection and justifying non-disclosure through fear of penalisation, peer judgement, or the belief that GenAI use is equivalent to using other accepted tools (e.g., Grammarly). In response, teachers reported adapting their practices through assignment redesign and trust-based classroom approaches rather than relying solely on detection.

The emotional impact of chatbots has also been studied in health and well-being contexts. The use of the SnehAI chatbot in India [S34] demonstrated that AI can create a safe and non-judgemental space for young people to learn about sexual and reproductive health, topics that are often taboo in their communities. This access to safe information directly influences behavioural decision-making and young people's autonomy over their own bodies. The authors [S28] reported that users described chatbots as an "emotional sanctuary," helping them deal with difficult moments but also expressing frustration about the limitations in empathic listening and the responses offered by the chatbots. This is complemented by information from another article [S29], showing that although systems offer support, they restrict epistemic activities essential to therapy, such as self-understanding and critical questioning. Therefore, even though there is evidence of immediate emotional benefit, studies warn that the support offered by chatbots does not fully replace the depth of human therapeutic interaction. A large-scale Reddit analysis of conversations with ChatGPT [S53] corroborates this pattern: users favoured ChatGPT over human support for its constant availability, non-judgemental stance, and privacy, using it for actionable advice, emotional validation, and sensemaking of confusing thoughts, while also warning of risks such as inaccurate mental health information, ChatGPT's tendency to affirm harmful beliefs uncritically, and privacy concerns when disclosing sensitive details.

To conclude the analysis, two studies [S19, S25] address how individual psychological characteristics influence the use of AI technologies. One [S19] indicates that users with low self-control and low self-efficacy tend to use AI tools more frequently, especially for academic tasks, while those with higher self-esteem tend to use these systems less. The study by [S25] demonstrates that the acceptance and perceived usefulness of technology are associated with reduced anxiety among students, suggesting that emotional and cognitive factors modulate not only the use but also the subjective impacts of these tools. Both studies reveal how the effects of chatbots are deeply linked to the psychological profiles of users and are neither uniform nor universal.

5.4.1 RQ4.1: Within these impacts, what evidence regarding users with physical or cognitive difficulties is reported?

To complement RQ4, this subsection synthesises whether the included studies explicitly report chatbot use by individuals with physical or cognitive difficulties. Here, we use "evidence" to denote direct reporting in the reviewed studies. We also summarise inclusion or support-related considerations discussed in these studies when present.

Study [S48] provides the first empirical evidence in this review of chatbot use among students with physical disabilities (including Blind/Low-Vision, Deaf/Hard of Hearing, and Mobility or Dexterity disabilities), reporting benefits such as reformatting inaccessible materials, drafting accommodation-related communications, and reducing reliance on formal disability support services. One further article [S56] complements this from a research-design perspective. Rather than reporting empirical findings, S56 presents a PhD research proposal that explicitly targets the intersection between generative AI and accessibility for people with physical disabilities. The author argues that current chatbot interaction remains dominated by chat-based input windows, which are not optimised for users who rely on assistive technologies, and proposes deformable interfaces, tangible artefacts that support alternative interaction modalities such as bending, stretching, or pulling, as a means of diversifying how people with disabilities can engage with LLMs. While S56 does not yet provide the empirical data our review sought, it signals a growing research interest in multimodal, inclusive AI interaction and reinforces the gap identified in this review: the near-total absence of empirical studies evaluating chatbot or AI agent use among people with physical disabilities. It is worth noting that study [S43], although not designed to investigate disability, incidentally included older adult participants with visual, hearing, and mobility impairments among its five lab-study participants, offering limited but relevant anecdotal context on how such users interact with LLM-powered conversational companions.

Most of the articles do not directly mention the impact of these tools on people with cognitive complications. However, eight studies addressed this topic [S9, S15, S20, S25, S26, S28, S48, S53]. We noticed that articles focusing more on mental health [S25, S26, S28] were better at answering our RQ, while articles that focused more on the subject of user behaviour did not address the topic of the question as much.

The number of individuals suffering from mental disorders has grown significantly in our society in recent decades. The growth in these cases rose 48% from 1990 to 2019, reaching 980 million people globally [S28]. Therefore, it is important to evaluate the evidence that this instrument can be used as support for people with mental problems. In the domain of humanised interactions, one study [S15] argues that social robots and CAs can play a crucial role in improving the quality of life for people by mitigating behavioural biases and improving investment decision-making outcomes. These tools are often seen as a means of helping with daily tasks, but their abilities are often underestimated. These tools can play an important role as emotional support, bringing empathy and reassurance to some users. This functionality of chatbots is found in another study [S20], which emphasises

this "human" interaction with the user, but in an educational context. While the first [S15] focuses broadly on emotional well-being, the other [S20] narrows its scope to classroom engagement, highlighting how bots that respond adaptively to user feedback can increase participation among students with emotional problems. The study addresses the ability of these agents to listen and empathise with the user, helping students with cognitive complications.

Addressing the topic of humanising chatbots, a study [S9] shows that users with strong emotional reactions tend to be more resistant to using chatbots and also have a worse UX when faced with a system failure. According to the authors, these users tend to believe that a chatbot cannot help them with their tasks and refuse to use it. However, these same users also demonstrate a greater propensity to seek meaningful connections with bots as a way to deal with their emotional vulnerabilities. Another article [S15] argues that social robots and CAs can be important allies in improving the quality of life for people with mental disabilities, especially when integrated into controlled environments. The study argues that these technologies can offer both practical (reminders, simple instructions) and emotional support, especially when the systems adopt empathetic dialogue strategies.

Two studies have revealed the benefits of using AI technologies; one [S25] delved deeper into mental illnesses (such as depression and anxiety) and how this type of technology can help address these problems. The article states that chatbots can be strong allies for these individuals in several aspects, indicating that even in sensitive contexts such as mental health, chatbots can reduce communication barriers for people with social issues. It also emphasises the benefits of mobile applications made specifically for this audience but warns about the care that companies must take in the process and implementation of these tools. The other study [S26] comments that AI technology can be a strong emotional support for its users, overcoming the communication barrier and becoming a companion, especially for lonely people who need emotional support and often struggle with cognitive difficulties related to depression. This further emphasises the conclusion of the previously cited article, showing how these systems promote a more accessible and welcoming space for speech.

Complementing these findings, the study [S28] specifically investigates two popular mental health chatbots (Woebot and Wysa), highlighting their use by individuals with psychological needs. The results reveal that, although the therapeutic effect of this type of tool is small and does not last over time, this procedure in the search for psychological support can cause frustration for users, bringing feelings of emptiness, generality, nonsensicality, repetitiveness, and constraint.

Finally, study [S48] reported that students with disabilities, particularly those with ADHD, autism, and psychological/psychiatric disabilities, used GenAI tools to personalise information processing, draft accommodation-related emails, and reduce anxiety associated with self-advocacy. Participants also noted concerns regarding academic integrity suspicions, unclear course AI policies, and the absence of disability-specific guidance from instructors, reinforcing the need for accessibility-aware AI policy in higher education.

5.5 RQ5: How are human factors and ethical considerations addressed in the analysed studies?

In RQ5, we will explore how human and ethical aspects are addressed in the design of chatbots within the 56 studies reviewed. From the HCI perspective, ethical and human concerns are directly related to the interaction design choices, such as chatbot communication, representation, user adaptation, and the influence on users' behaviour. These studies delve into a wide range of diverse topics, leading to nine categories based on the kind of ethical aspects each study addressed: Research Practices, Injustices in AI, Chatbot Trust & Human-likeness, Education & Professional, UX & Humane AI, AI in Health, Personality Adaptation & Persuasion, Unethical Behaviour in AI, and Responsible AI Bias & Privacy. While ethical considerations are present throughout our research, we will now focus on how ethical aspects are treated within the human-AI interaction context, highlighting specific ethical concerns that differentiate RQ5 from the other RQs. The categories show how ethical considerations are embedded within design decisions, evaluations, and usage contexts.

The Research Practices category includes sixteen studies [S1, S6, S19, S25, S26, S28, S30, S32, S33, S34, S38, S39, S45, S46, S47, S49], all prioritising participant rights, privacy, and safety through measures such as ethical approval, anonymisation, and robust data protection, even without explicitly detailing human factors. One study [S39] obtained approval from the Brazilian National Commission for Research Ethics and collected individual informed consent. Similarly, three studies [S34, S38, S32] reported formal ethics committee approval and written consent procedures, while another study [S33] described institutional approval procedures. One study [S47] formalised data governance in multi-national and multi-institutional contexts.

The Injustices in the AI category consist of 2 subcategories: Bias & Epistemic Injustice in AI Data and Participatory Injustice in AI, each addressing distinct kinds of injustices within human-AI interaction. The Bias & Epistemic Injustice in AI Data subcategory covers two studies [S11, S29] that highlight ethical concerns about human participation in data analysis, interpretation, and sharing. They warn of growing design biases that impact the way concerns associated with gender, race, age, income inequality, social status, etc., are treated, stressing the importance of addressing issues related to authenticity, autonomy, and epistemic fairness, especially for vulnerable populations. Lastly, in Participatory Injustice in AI, a study [S29] discusses epistemic injustice, specifically participatory injustice. They argue against using chatbots to resolve a problem in terms of distributive justice at the cost of causing new issues in terms of epistemic injustice. Another study [S56] emphasises that ethical AI design must be inclusive by default, moving away from conventional text-based chat to multimodal interfaces that accommodate diverse physical and cognitive abilities.

Chatbot Trust & Human-likeness contains four studies [S12, S27,S36, S54] that examine user trust in chatbots, noting how features like empathy and natural language can enhance credibility despite initial distrust due to unmet expectations. One study [S36] addressed human-likeness directly by de-

signing a humanised doctor persona, intentionally shifting tone depending on user needs. Another study [S54] identifies "perceived humanness" as a double-edged sword; although it reduces psychological distance, it needs clearer regulatory frameworks for the disclosure of autonomy.

Education & Professional encompasses four studies [S25, S40, S48, S51], which delve into ethical concerns in educational and professional settings, such as academic integrity, job displacement, student anxiety, the stigma of using AI, transparency, and balancing AI benefits with issues of trust, privacy, moral values, and responsible behaviour.

UX & Humane AI explicitly addresses UX as a human factor. Three studies [S5,S9,S37] found that more experienced users perceive a stronger social presence, highlighting the need to design systems for both novice and experienced users. One study [S5] also noted the ethical importance of fostering a humane and user-friendly environment to encourage user adoption. Another study [S37] emphasised humane interaction design through a co-design approach, described as being designed "with and for young people." [Ludin *et al.*, 2022]

The AI in health category contains two studies [S50, S53] highlighting the dangers of AI hallucinations and the risk of reinforcing cognitive distortions when it provides overly affirmative or inaccurate health advices.

Personality Adaptation & Persuasion features one study [S4] that explores the ethical implications of using personality adaptation in chatbots to improve UX and trust, emphasising the need for strict guidelines to mitigate manipulative effects.

Unethical Behaviour in AI includes one study [S11], which showed how the absence of human presence in AI interactions can lead users to prioritise self-interest over moral standards, underscoring the importance of human representation in fostering ethical conduct.

The Responsible AI: Bias & Privacy category did not mention human aspects but highlighted key ethical concerns such as bias, data security, and privacy, underscoring the importance of responsible AI development. [S2]

In Table 8, we can visualise all categories along with their subcategories, objectives, and articles. In Figure 8, we can see the number of articles in each category for RQ5.

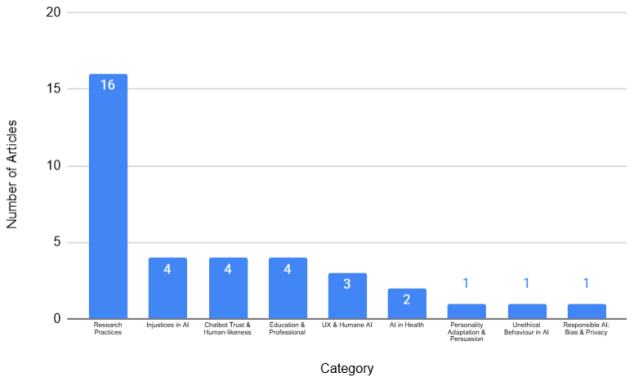

Figure 8. Number of articles per category.

6 Discussion

In this subsection, we will outline the answers for each RQ described in the previous section.

From our findings in RQ1, we noticed that ChatGPT was the chatbot used in most of the approaches, while the contexts varied across several aspects; however, many were placed in specific environments, such as education, mental health, trust, and consumption. We also noticed that the type of chatbot may be related to its functionality; chatbots aimed at consumption, for example, prioritise simulating efficient human service, thereby bringing trust to the user, while educational chatbots aim to promote learning.

Moving to RQ2, we found very different purposes for each study, such as the ethical Aspects, chatbots in education, chatbots in healthcare, chatbots in customer experience, chatbot avatars and user perception, enhancing AI user interaction, and explainable chatbots. These results show that although their purposes are varied, they follow a similar path, which is to evaluate the use of chatbots in different contexts. The importance of these objectives is highlighted in the way the studies demonstrate that researchers and developers are collaborating to design chatbot functionalities and ethical considerations to meet specific requirements in each chatbot's context of use, considering the results from the evaluation procedures.

For RQ3, we noticed that the interaction, as well as how it was perceived, differed depending on the context of use. From an ethical perspective, the interaction often centred on functional tasks. Users were more likely to lose trust and less likely to continue using these chatbots after experiencing a breakdown, especially when engaged in complex tasks. We also had positive user perceptions despite concerns such as privacy and technological immaturity. In AI in Health contexts, interactions were described as more sensitive and emotionally loaded, focused on natural communication, empathy, questioning behaviour, and perceived human-likeness, especially in cases where chatbots were designed for counselling, mental health support, symptom checking, or vulnerable populations. In customer experience, the interaction was more focused on online shopping, assessing how chatbot interactions influence customer satisfaction and behaviour, with no detailed user perceptions. In the educational context, ChatGPT was widely used to explore how AI affects students' learning experiences, showing student vulnerability due to a lack of training and increasing concerns about safety and privacy with the widespread adoption. To enhance AI-user interaction and chatbot avatars, we reviewed interactions from different systems, as well as pre-interaction expectations and post-interaction feedback. From user perception, human-like qualities may lead to higher user acceptance and foster emotional bonds, especially for lonely individuals seeking social connection. When using AI for data queries and insights, the interactions occurred specifically within the domain of log management and security monitoring, highlighting the need to enable users to efficiently search, filter, and analyse log data. Lastly, for participatory injustice, the interaction context was in application development, where individuals were treated merely as "testing subjects." We noticed that when users' contributions are not taken seriously, it can lead to individuals

Table 8. RQ5 - Categories, Subcategories, Objectives, and Articles

Category	Subcategories	Objective	Articles
Research Practices	–	Ensure ethical handling of participants' data through anonymization, consent, approval, and safety measures, even without explicitly detailing human factors.	S1 [Fabio *et al*., 2024], S6 [Mubassira *et al*., 2025], S19 [Rodríguez-Ruiz *et al*., 2025], S25 [Yeh *et al*., 2025], S26 [Peng *et al*., 2024], S28 [Siddals *et al*., 2024], S30 [Maples *et al*., 2024], S32 [Kang and Hong, 2025], S33 [Shan *et al*., 2022], S34 [He *et al*., 2022], S38 [Wong *et al*., 2021], S39 [Chagas *et al*., 2023], S45 [Wang *et al*., 2025], S46 [He *et al*., 2025], S47 [Marmolejo-Ramos *et al*., 2026], S49 [Kang *et al*., 2025]
Injustices in AI	Bias & Epistemic Injustice in AI Data	Address epistemic fairness, authenticity, and representation issues in AI data involving human demographics such as race, gender, and income, especially for vulnerable populations.	S11 [Kim *et al*., 2023], S29 [Pozzi and De Proost, 2024]
	Participatory Injustice in AI	Reveal how the use of chatbots to resolve distributive justice can cause epistemic injustice by excluding certain populations from participatory knowledge processes.	S29 [Pozzi and De Proost, 2024], S56 [Thalhammer, 2025]
Chatbot Trust & Human-likeness	–	Explore how chatbot empathy and natural language can increase user trust, despite initial distrust due to unmet expectations.	S12 [Saoud and Romman, 2024], S27 [Park *et al*., 2024], S36 [Chung *et al*., 2021], S54 [Hasselwander and Lah, 2026]
Education & Professional	–	Investigate ethical challenges in educational and professional settings, including trust, privacy, academic integrity, and job displacement.	S25 [Yeh *et al*., 2025], S40 [Adnin *et al*., 2025], S48 [Atcheson *et al*., 2025], S51 [Alshammari *et al*., 2025]
UX & Humane AI	–	Explicitly addresses UX as a human factor, highlighting the ethical value of UX, showing how system design can enhance social presence and inclusivity for both novice and experienced users.	S5 [Meng *et al*., 2024], S9 [Min *et al*., 2021], S37 [Ludin *et al*., 2022]
AI in Health	–	Address hallucinations, cognitive reinforcement, and privacy risks in mental health support.	S50 [Mayerhofer *et al*., 2025], S53 [Jung *et al*., 2025]
Personality Adaptation & Persuasion	–	Assess ethical implications of chatbots adapting their personality to users, highlighting risks of manipulation and the need for regulation.	S4 [Joshi and Bengler, 2024]
Unethical Behaviour in AI	–	Examine how lack of human presence in AI systems may lead users to unethical decisions, stressing the importance of human representation.	S11 [Kim *et al*., 2023]
Responsible AI: Bias & Privacy	–	Emphasize the ethical need for responsible AI development by addressing concerns related to bias, data privacy, and security, though without focusing on human aspects directly.	S2 [Akhtar *et al*., 2024]

being disempowered, raising mental health concerns.

Regarding the user experience rating in the system, we found distinct types for rating UX in the reviewed articles, such as feedback, data collection, and analysis techniques (five-point Likert scale in an online survey), statistical tests, qualitative analysis, sampling designs, demographic analysis, mixed-method user evaluation, and quantitative assessments (hypothesis testing, statistical analysis, descriptive statistics, and multiple linear regression). We also noticed that many studies combined different methodologies to gain a comprehensive understanding of UX, which revealed mixed user opinions. Additionally, the studies that used experimental designs and diverse samples are notable for their inclusion of diverse and sometimes vulnerable populations, providing crucial insights into real-world scenarios. However, it is also important to mention that some studies did not provide evidence of how the evaluation was carried out, and many did not report participant details, being limited by their UX findings.

Based on the findings of RQ4, the use of chatbots has demonstrated impacts on several dimensions of user's lives, such as loneliness, educational engagement, support at work, and even ethical behaviour. Most articles present these tools as beneficial to their users, evidencing improvements in critical thinking, student engagement, and emotional support. However, risks were highlighted, such as dependence on and replacement of human relationships by AI. We did not find any studies that addressed the long-term impacts of chatbots on users' lives, mainly because AI was created recently and has evolved very quickly over the years.

Regarding RQ4.1, no empirical evidence was found concerning assistance for people with physical disabilities; only a research proposal [S56] outlining multimodal, accessibility-focused interfaces, which indicates a possible gap to explore in future empirical studies. As for cognitive difficulties, many studies have shown benefits, especially in mental health contexts. Chatbots have shown potential to combat anxiety and isolation, but they have certain limitations. Due to their quick, constant, and inexpensive access, many people seek these tools for emotional support; however, this relationship is superficial and can generate frustration for the user.

In RQ5, we found many different ways in which both human factors and ethical aspects are treated; for example, we have studies focused on research practices, prioritising participant rights, privacy, and safety through measures such as ethical approval, anonymisation, and robust data protection, as well as others highlighting ethical concerns related to human participation, such as gender, race, age, income inequality, social status, bias, and data security. In terms of human trust in chatbots, we noticed that features like empathy and natural language can enhance credibility. In the educational context, ethical concerns such as academic integrity and job displacement were highlighted. Considering chatbots with Personality Adaptation & Persuasion features, the ethical implications of using personality adaptation in chatbots to improve UX and trust were explored. In terms of AI behaviour, the absence of human presence in AI interactions can lead users to prioritise self-interest over moral standards, meaning that human representation is important for ethical conduct.

The RQs showed that the interaction between chatbots and humans cannot be understood in isolation. The type and context of the application of each chatbot directly shape the desired objectives as well as the UX. The development of these tools is not only about efficiency but also about enabling and providing a human relationship that respects human particularities and their complexities.

The predominance of generic chatbots, especially those based on LLMs such as ChatGPT, highlights a demand for more versatile tools that are accessible to different needs. Although this objective is increasingly becoming a reality, it presents some risks. A machine designed for a very large number of people who have different economic, social, and regional realities must be developed meticulously. This problem is evidenced by the lack of studies that address the use of chatbots by people with physical disabilities. However, the strong presence of research studies discussing their use within the context of mental health demonstrates that there is attention to the development of tools that do not cause or aggravate negative feelings, such as loneliness.

It is important that users do not become frustrated by creating high expectations for chatbots and not receiving the expected responses. Some studies have addressed issues related to people's trust in these tools, highlighting concerns that are more intense in sensitive contexts such as health and education. Aspects such as appearance, clarity, and transparency have proven to be essential for good UX, demonstrating the high subjectivity of this relationship.

The impact of chatbots goes beyond their immediate functionality, influencing people's decisions and ethical behaviours. These systems can become "emotional sanctuaries," helping people with emotional problems. At the same time, warnings are being issued, especially regarding excessive use, which can cause dependency and replace human relationships. Other problems, such as cognitive laziness and loss of thinking skills, are also real. However, some articles highlight the stimulation of creativity and student engagement. All of these statements may seem contradictory, which only demonstrates the complexity of chatbots and how their impact is not uniform, varying according to the intentionality, context, and profile of each user.

It is important that chatbots are not created only for their technical performance; their sensitivity to dialogue with human needs must be considered an important factor. To advance in this field, technical studies that incorporate an ethics of digital care, which considers the development of these systems based on justice, inclusion, and empathy, are necessary. Users from diverse backgrounds must be considered at the beginning of any project of this kind, taking into account their real experiences and living conditions. This implies the ethics of development, which is one of the most important aspects to consider when creating AI tools. It lies in the transparency of the treatment of sensitive data, the choice of the audience to be considered in the project, the human representation that communicates, and others.

6.1 Limitations of the SLR and threats to validity

In this section, we show the limitations of our SLR and the associated threats to validity (construct, internal, external, and conclusion), openly discussing the impact of our methodological decisions.

In terms of construct validity, the RQs and coding categories may not adequately capture human factors, UX, and ethical aspects, leaving some dimensions unrepresented, such as cultural differences and cognitive effects. Some ethical considerations might have been left unattended as well since the studies were categorised based on the most explicit aspect discussed within each one.

For internal validity, there may be a possible misclassification of categories. Since the categorisation of the studies was conducted by only two main screeners, the thematic grouping involved some degree of subjectivity, which could lead to potential misclassification. Although disagreements were resolved through consultation with a third reviewer, thereby reducing individual bias, it did not entirely eliminate the risk of subjective interpretation.

In addition, because individual pre-consensus screening decisions were not preserved in a structured format, no inter-rater reliability measure could be calculated. This limits the quantitative assessment of agreement between reviewers, although independent screening and consensus discussion were used to reduce selection bias.

Moving to external validity, this review was restricted to the ACM Digital Library, IEEE Xplore, SpringerLink, ScienceDirect, and Web of Science, which were selected for their coverage of Computing, HCI, AI-based interactive systems, and related interdisciplinary areas. Although these sources cover relevant literature in the field, some studies indexed only in other relevant databases for Computing and HCI, such as Scopus, may have been missed. This may affect the completeness of the corpus and the generalisability of the findings. In addition, studies from related areas, such as psychology and education, may have been excluded if they were not indexed in the selected sources. The review was also limited to studies published in English and to the period between January 2020 and January 2026. Although this time frame was motivated by the rapid evolution of conversational AI and recent forms of human–chatbot interaction, it may have excluded earlier foundational work on human–chatbot interaction and UX.

Lastly, regarding conclusion validity, our study did not conduct a formal quality assessment of primary studies, which might affect how strongly conclusions can be generalised. All included studies were considered equally informative during the synthesis, which may have allowed less robust studies to influence the overall conclusions. As a result, our conclusion should be interpreted as indicative patterns and research trends rather than definitive claims.

7 Conclusion

In this section, we will present the limitations and possible future research studies that could be added to our study.

This SLR aimed to understand how AI-based chatbots have been used, perceived, and evaluated by their users and, thus, to understand the challenges and possible solutions that this issue entails. We paid special attention to the human, ethical, and behavioural factors that involve chatbots and their users. Through the analysis of 56 studies, we identified divisions based on the context in which they are situated, namely

behaviour, consumer, education, and mental health. There was a predominance of the use of generic and conversational chatbots, with an emphasis on ChatGPT. The investigations revealed that UX is highly influenced by variables such as empathy, trust, usability, and application context.

The results of our analysis demonstrate a positive impact regarding chatbots. Although their impact is not uniform and risk-free, these tools have great potential to assist in daily activities and promote well-being and emotional support. However, some cons were observed. Their excessive use can generate problems such as frustration, dependence, and social isolation. Nonetheless, the research highlights positive factors, given the accessibility and support offered by these systems.

Ethical issues were also the focus of our review. Some studies addressed concerns such as fairness, transparency, and consent, demonstrating an important area for improvement and clarification by AI companies. Ethical responsibility must be taken seriously and must keep pace with the technological evolution of these tools, especially given their strong contact with human beings.

Based on our entire analysis, we realised that some aspects have the potential to be further explored in the future. Due to the recent creation of AI-based chatbots, we did not find any studies that analysed their long-term impact. It is extremely important that other research studies can be conducted considering a long time span since this scenario has not been studied before, especially given the potential impact of AI on the daily lives of its users. We missed empirical studies that addressed the use of these tools for individuals with physical limitations, though one design-orientated proposal [S56] points in this direction. We observed a strong debate regarding the mental health of users, but the possibilities of helping people with physical disabilities are significant and cannot be overlooked. Additionally, we noticed that the reviewed studies did not explore UI and UX aspects in detail, only going so far as to explore generic user perception; this would be a great start for future research studies to delve deeper into the UI and UX domain.

As a recommendation for future research, ethical and human aspects should be observed in close detail, as people tend to become very dependent on chatbots' help. Key areas of focus should include aspects of the human-chatbot interaction, such as security, transparency, and the truthfulness of chatbot-generated information, accounting for misinformation and potential harm.

7.1 Practical implications

Our findings suggest several practical implications for different stakeholders with a direct connection to chatbots.

7.1.1 Implications for chatbot designers and UX professionals

The main implications for chatbot designers and UX professionals emphasised by our findings are the importance of empathy and transparency. Features such as emotionally aware responses and clear communication of the system can positively influence trust and user satisfaction. However, over-anthropomorphism in sensitive domains, such as mental health, may create emotional dependency and should be avoided. Chatbot interactions should convey system limitations and be goal-orientated.

7.1.2 Implications for educators and institutions

In the educational context, our findings demonstrate that chatbots show potential as tools for learning support, feedback, and productivity. However, educational institutions should be able to balance the support offered by the chatbots with the risk of dependency. The reviewed studies suggest a need for training students and educators in the critical use of AI tools, with clear institutional guidelines that help ensure chatbots are used to complement learning rather than replace cognitive and pedagogical actions.

7.1.3 Implications for mental health and vulnerable populations

Considering mental health and vulnerable populations, our findings show that chatbots can provide support for mental health and well-being. However, our findings also highlight the risks of frustration related to shallow empathy; therefore, chatbots should be used as complementary support and not as a substitute for professional help, encouraging the participation of human professionals. Additionally, special attention should be given to vulnerable populations to ensure that chatbot interactions are safe, transparent, and ethically responsible.

7.1.4 Implications for policy and ethics

Ethical concerns prevail across our reviewed studies, highlighting the need for guidelines on fairness, privacy, and long-term effects that are important to support responsible AI development. Policymakers and organisations should promote ethical standards that explicitly consider human–AI interaction and user well-being, ensuring that chatbot technologies align with social values and are designed in a human-centred way.

Declarations

Acknowledgements

Lívia Lutz & Breno Lois Matos were supported by undergraduate research scholarships from the Brazilian National Council for Scientific and Technological Development (CNPq) through the Institutional Program for Scientific Initiation Scholarships (PIBIC).

Authors' Contributions

Lívia Lutz & Breno Lois Matos: Conceptualisation, Data curation, Formal analysis, Investigation, Visualisation, Writing – original draft. Greis Francy M. Silva-Calpa: Conceptualisation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The dataset generated and analyzed during the current study is fully available in the Zenodo repository. This repository contains the exported screening records and the decisions applied during title, abstract, and full-text screening and can be accessed via at: `https://doi.org/10.5281/zenodo.21146843`.

References

Adamopoulou, E. and Moussiades, L. (2020a). Chatbots: History, technology, and applications. *Machine Learning with applications*, 2:100006. DOI: https://doi.org/10.1016/j.mlwa.2020.100006.

Adamopoulou, E. and Moussiades, L. (2020b). An overview of chatbot technology. In Maglogiannis, I., Iliadis, L., and Pimenidis, E., editors, *Artificial Intelligence Applications and Innovations*, pages 373–383, Cham. Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-49161-8_31.

Adnin, R., Pandkar, A., Yao, B., Wang, D., and Das, M. (2025). Examining student and teacher perspectives on undisclosed use of generative ai in academic work. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3706598.3713393.

Akhtar, M. S., Mathur, G., Rakhra, M., and Sharma, C. (2024). Ai bots: Navigating the next wave of innovation - 2024 international conference on electrical electronics and computing technologies (iceect). 1:1–7. DOI: https://doi.org/10.1109/ICEECT61758.2024.10738911.

Alharbi, N., Ud Din, F., Paul, D., and Sadgrove, E. (2025). Driving ai chatbot adoption: A systematic review of factors, barriers, and future research directions. *Journal of Open Innovation: Technology, Market, and Complexity*, 11(3):100590. DOI: https://doi.org/10.1016/j.joitmc.2025.100590.

Alshammari, S. H., Alrashidi, M. E., Alshammari, M. H., Alshammari, A. E. A., and Alkhwaldi, A. F. (2025). Determinants of student adoption of artificial intelligence applications in higher education. *Scientific Reports*, 15(1):35921. DOI: https://doi.org/10.1038/s41598-025-19851-5.

Alshammari, S. H. and Babu, E. (2025). The mediating role of satisfaction in the relationship between perceived usefulness, perceived ease of use and students' behavioural intention to use chatgpt. *Scientific Reports*, 15(1):7169. DOI: https://doi.org/10.1038/s41598-025-91634-4.

Atcheson, A., Khan, O., Siemann, B., Jain, A., and Karahalios, K. (2025). "i'd never actually realized how big an impact it had until now": Perspectives of university students with disabilities on generative artificial intelligence. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3706598.3714121.

Aumüller, A., Winklbauer, A., Schreibmaier, B., Batinic, B., and Mara, M. (2024). Rethinking feminized service bots: user responses to abstract and gender-ambiguous chatbot avatars in a large-scale interaction study. *Personal and Ubiquitous Computing*, 28(6):1021–1032. DOI: https://doi.org/10.1007/s00779-024-01830-8.

Balaji, K. and Rao, P. (2024). Artificial intelligence and chatbots: Transforming user experience in e-tailing. pages 1375–1379. DOI: https://doi.org/10.1109/ICCES63552.2024.10859474.

Belda-Medina, J. and Calvo-Ferrer, J. R. (2022). Using chatbots as ai conversational partners in language learning. *Applied Sciences*, 12(17):8427. DOI: https://doi.org/10.3390/app12178427.

Caldarini, G., Jaf, S., and McGarry, K. (2022). A literature survey of recent advances in chatbots. *Information*, 13(1):41. DOI: https://doi.org/10.3390/info13010041.

Chagas, B. A., Pagano, A. S., Prates, R. O., Praes, E. C., Ferreguetti, K., Vaz, H., Nogueira Reis, Z. S., Ribeiro, L. B., Ribeiro, A. L. P., Pedroso, T. M., Beleigoli, A., Oliveira, C. R. A., and Marcolino, M. S. (2023). Evaluating user experience with a chatbot designed as a public health response to the covid-19 pandemic in brazil: Mixed methods study. *JMIR Human Factors*, 10(1):e43135. DOI: https://doi.org/10.2196/43135.

Chung, K., Cho, H. Y., and Park, J. Y. (2021). A chatbot for perinatal women's and partners' obstetric and mental health care: Development and usability evaluation study. *JMIR Medical Informatics*, 9(3):e18607. DOI: https://doi.org/10.2196/18607.

Dempere, J., Modugu, K., Hesham, A., and Ramasamy, L. K. (2023). The impact of chatgpt on higher education. In *Frontiers in Education*, volume 8, page 1206936. Frontiers Media SA. DOI: https://doi.org/10.3389/feduc.2023.1206936.

Deng, Y., Zhao, N., and Huang, X. (2023). Early chatgpt user portrait through the lens of data. In *2023 IEEE International Conference on Big Data (BigData)*, pages 4770–4775. DOI: https://doi.org/10.1109/BigData59044.2023.10386415.

Ding, Y. and Najaf, M. (2024). Interactivity, humanness, and trust: a psychological approach to AI chatbot adoption in e-commerce. *BMC Psychology*, 12(1):595. DOI: https://doi.org/10.1186/s40359-024-02083-z.

Fabio, R. A., Plebe, A., and Suriano, R. (2024). AI-based chatbot interactions and critical thinking skills: an exploratory study. *Current Psychology*. DOI: https://doi.org/10.1007/s12144-024-06795-8.

Genç, H. U., Chandrasegaran, S., Dingler, T., and Verma, H. (2025). Persuasion in pixels and prose: The effects of emotional language and visuals in agent conversations on decision-making. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. DOI: https://doi.org/10.1145/3706598.3713579.

Guerino, G. C. and Valentim, N. M. C. (2020). Usability and user experience evaluation of conversational systems: A systematic mapping study. In *Proceedings of the 34th Brazilian Symposium on Software Engineering (SBES '20)*, pages 427–436. DOI: https://doi.org/10.1145/3422392.3422421.

Gumusel, E. (2025). A literature review of user privacy concerns in conversational chatbots: A social informatics approach. *Journal of the Association for Information Science and Technology*, 76(1):121–154. DOI: https://doi.org/10.1002/asi.24898.

Hasselwander, M. and Lah, O. (2026). Agentic ai arrives: How gen z adopts autonomous ai agents. *IEEE Access*, 14:27083–27090. DOI: https://doi.org/10.1109/ACCESS.2026.3665348.

Haugeland, I. K. F., Følstad, A., Taylor, C., and Bjørkli, C. A. (2022). Understanding the user experience

of customer service chatbots: An experimental study of chatbot interaction design. *International Journal of Human–Computer Studies*, 161:102788. DOI: https://doi.org/10.1016/j.ijhcs.2022.102788.

He, G., Aishwarya, N., and Gadiraju, U. (2025). Is conversational xai all you need? human-ai decision making with a conversational xai assistant. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*, IUI '25, page 907–924, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3708359.3712133.

He, Y., Yang, L., Zhu, X., Wu, B., Zhang, S., Qian, C., and Tian, T. (2022). Mental health chatbot for young adults with depressive symptoms during the covid-19 pandemic: Single-blind, three-arm randomized controlled trial. *Journal of Medical Internet Research*, 24(11):e40719. DOI: https://doi.org/10.2196/40719.

Hirschberg, J. and Manning, C. D. (2015). Advances in natural language processing. *Science*, 349(6245):261–266. DOI: https://doi.org/10.1126/science.aaa8685.

Iberdrola (2025). What is a chatbot? `https://www.iberdrola.com/innovation/what-is-a-chatbot`. Accessed: 23 August 2026.

Jain, V., Singh, I., Syed, M., Mondal, S., and Ranjan Palai, D. (2024). Enhancing educational interactions: A comprehensive review of ai chatbots in learning environments - 2024 11th international conference on reliability, infocom technologies and optimization (trends and future directions) (icrito). pages 1–5. DOI: https://doi.org/10.1109/ICRITO61523.2024.10522392.

Joshi, R. and Bengler, K. (2024). Crafting human-ai interaction: A rhetorical approach to adaptive interaction in conversational agents - proceedings of the 12th international conference on human-agent interaction. page 314–322. DOI: https://doi.org/10.1145/3687272.3688297.

Jung, K., Lee, G., Huang, Y., and Chen, Y. (2025). 'i've talked to chatgpt about my issues last night.': Examining mental health conversations with large language models through reddit analysis. *Proceedings of the ACM on Human-Computer Interaction*, 9(CSCW3):1–25. DOI: https://doi.org/10.1145/3757537.

Kang, B. and Hong, M. (2025). Development and evaluation of a mental health chatbot using chatgpt 4.0: Mixed methods user experience study with korean users. *JMIR Medical Informatics*, 13:e63538. DOI: https://doi.org/10.2196/63538.

Kang, S., Potinteu, A.-E., and Said, N. (2025). Explainitai: When do we trust artificial intelligence? the influence of content and explainability in a cross-cultural comparison. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '25, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3706599.3720222.

Karataş, F., Abedi, F. Y., Ozek Gunyel, F., Karadeniz, D., and Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*, 29(15):19343–19366. DOI: https://doi.org/10.1007/s10639-024-12574-6.

Khurana, A., Alamzadeh, P., and Chilana, P. K. (2021). Chatrex: Designing explainable chatbot interfaces for enhancing usefulness, transparency, and trust - 2021 ieee symposium on visual languages and human-centric computing (vl/hcc). pages 1–11. DOI: https://doi.org/10.1109/VL/HCC51201.2021.9576440.

Kim, T., Lee, H., Kim, M. Y., Kim, S., and Duhachek, A. (2023). AI increases unethical consumer behavior due to reduced anticipatory guilt. *Journal of the Academy of Marketing Science*, 51(4):785–801. DOI: https://doi.org/10.1007/s11747-021-00832-9.

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., and Linkman, S. (2009). Systematic literature reviews in software engineering–a systematic literature review. *Information and software technology*, 51(1):7–15. DOI: https://doi.org/10.1016/j.infsof.2008.09.009.

Kooli, C. (2023). Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability*, 15(7):5614. DOI: https://doi.org/10.3390/su15075614.

Kothari, D. K. and Noel Newton Fernando, O. (2024). Enhancing human-computer interaction through ai: A study on chatgpt in educational environments - 2024 ieee conference on artificial intelligence (cai). pages 500–503. DOI: https://doi.org/10.1109/CAI59869.2024.00100.

Limna, P., Kraiwanit, T., Jangjarat, K., Klayklung, P., and Chocksathaporn, P. (2023). The use of chatgpt in the digital era: Perspectives on chatbot implementation. *Journal of Applied Learning & Teaching*, 6(1):64–74. DOI: https://doi.org/10.37074/jalt.2023.6.1.32.

Liu, Y., Li, X., and Xiang, Z. (2022). The effect of chatbot-customer interaction on consumer brand advocacy: Exploring the role of chatbots - 2022 ieee 12th international conference on electronics information and emergency communication (iceiec). pages 185–190. DOI: https://doi.org/10.1109/ICEIEC54567.2022.9835050.

Ludin, N., Holt-Quick, C., Hopkins, S., Stasiak, K., Hetrick, S., Warren, J., and Cargo, T. (2022). A chatbot to support young people during the covid-19 pandemic in new zealand: Evaluation of the real-world rollout of an open trial. *Journal of Medical Internet Research*, 24(11):e38743. DOI: https://doi.org/10.2196/38743.

Luo, B., Lau, R. Y., Li, C., and Si, Y.-W. (2022). A critical review of state-of-the-art chatbot designs and applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(1):e1434. DOI: https://doi.org/10.1002/widm.1434.

Maples, B., Cerit, M., Vishwanath, A., and Pea, R. (2024). Loneliness and suicide mitigation for students using GPT3-enabled chatbots. *npj Mental Health Research*, 3(1):4. DOI: https://doi.org/10.1038/s44184-023-00047-6.

Mariano, P. A. d. L., de Souza, A. C. R., Guerino, G. C., Chaves, A. P., and Valentim, N. M. C. (2024). A systematic mapping study about technologies for hedonic aspects evaluation in text-based chatbots. *Journal on Interactive Systems*, 15(1):875–896. DOI: https://doi.org/10.5753/jis.2024.4350.

Marmolejo-Ramos, F., Abadia, R., Karakale, , Barrera-Causil, C., Männikkö, N., Ghorbani, B. D., Strzelecki, A., Nwizu,

S., Tavares, C., Castillo, M., Escajal, B., Rodriguez, P., Som, B., Ngo, V. G., Özdoğru, A. A., Ariyabuddhiphongs, K., Thongrin, S., and Tejada, J. (2026). University students' perceptions and adoptions of ai: a cross-national study. *Discover Artificial Intelligence*, 6(1):320. DOI: https://doi.org/10.1007/s44163-026-01042-4.

Maroengsit, W., Piyakulpinyo, T., Phonyiam, K., Pongnumkul, S., Chaovalit, P., and Theeramunkong, T. (2019). A survey on evaluation methods for chatbots. In *Proceedings of the 2019 7th International conference on information and education technology*, pages 111–119. DOI: https://doi.org/10.1145/3323771.3323797.

May, R. and Denecke, K. (2021). Security, privacy, and healthcare-related conversational agents: A scoping review. *Informatics for Health and Social Care*, 47(2):194–210. DOI: https://doi.org/10.1080/17538157.2021.1983578.

Mayerhofer, K., Capra, R., and Elsweiler, D. (2025). Blending queries and conversations: Understanding trust, verification, and system choice in search and chat interactions. In *Proceedings of the 2025 ACM SIGIR Conference on Human Information Interaction and Retrieval*, pages 168–178. DOI: https://doi.org/10.1145/3698204.3716454.

Mehrotra, S., Degachi, C., Vereschak, O., Jonker, C. M., and Tielman, M. L. (2024). A systematic review on fostering appropriate trust in human-ai interaction: Trends, opportunities and challenges. *ACM Journal on Responsible Computing*, 1(4):1–45. DOI: https://doi.org/10.1145/3696449.

Meng, J., Rheu, M. M., Zhang, Y., Dai, Y., and Peng, W. (2023). Mediated social support for distress reduction: Ai chatbots vs. human. *Proc. ACM Hum.-Comput. Interact.*, 7. DOI: https://doi.org/10.1145/3579505.

Meng, S., Dollahite, R., Kaur, A., Schell, P., Sharma, A., Ventre, G., Kranz, I., Nudelman, G., and Gerling, G. J. (2024). Developing design features to facilitate ai-assisted user interactions - 2024 systems and information engineering design symposium (sieds). pages 437–442. DOI: https://doi.org/10.1109/SIEDS61124.2024.10534665.

Merelo Guervos, J. J., Castillo Valdivieso, P. Å., Mora García, A. M., Barranco Expósito, F., Guillén Perales, A., and Tsivitanidou, O. (2023). Chatbots and messaging platforms in the classroom: An analysis from the teacher's perspective. *Education and Information Technologies*, 29:1903–1938. DOI: https://doi.org/10.1007/s10639-023-11703-x.

Min, F., Fang, Z., He, Y., and Xuan, J. (2021). Research on users' trust of chatbots driven by ai: An empirical analysis based on system factors and user characteristics - 2021 ieee international conference on consumer electronics and computer engineering (iccece). pages 55–58. DOI: https://doi.org/10.1109/ICCECE51280.2021.9342098.

Mubassira, T., Hasan, M., Noor, J., and Islam, A. B. M. A. A. (2025). Enhancing emobot: An in-depth analysis of user expectation and satisfaction in an emotion-aware chatbot - proceedings of the 11th international conference on networking, systems, and security. page 65–71. DOI: https://doi.org/10.1145/3704522.3704527.

Ng, S. W. T. and Zhang, R. (2025). Trust in ai chatbots: A systematic review. *Telematics and Informatics*, 97:102240. DOI: https://doi.org/10.1016/j.tele.2025.102240.

Oracle (2024). What is natural language processing (NLP)? `https://www.oracle.com/uk/artificial-intelligence/natural-language-processing/`. Accessed: 23 August 2026.

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., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., and Moher, D. (2021a). The prisma 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372. DOI: https://doi.org/10.1136/bmj.n71.

Page, M. J., Moher, D., 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., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., and McKenzie, J. E. (2021b). Prisma 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372. DOI: https://doi.org/10.1136/bmj.n160.

Park, D.-M., Jeong, S.-S., and Seo, Y.-S. (2022). Systematic review on chatbot techniques and applications. *Journal of Information Processing Systems*, 18(1):26–47. DOI: https://doi.org/10.3745/JIPS.02.0175.

Park, G., Chung, J., and Lee, S. (2024). Human vs. machine-like representation in chatbot mental health counseling: the serial mediation of psychological distance and trust on compliance intention. *Current Psychology*, 43(5):4352–4363. DOI: https://doi.org/10.1007/s12144-023-04653-7.

Peng, C., Zhang, S., Wen, F., and Liu, K. (2024). How loneliness leads to the conversational AI usage intention: the roles of anthropomorphic interface, para-social interaction. *Current Psychology*. DOI: https://doi.org/10.1007/s12144-024-06809-5.

Pozzi, G. and De Proost, M. (2024). Keeping an AI on the mental health of vulnerable populations: reflections on the potential for participatory injustice. *AI and Ethics*. DOI: https://doi.org/10.1007/s43681-024-00523-5.

Putra Antonio, E. R., Fadhilah, M. F., Faiq, F., Fredyan, R., and Pranoto, H. (2023). Analyzing the impact of customer service chatbots on user satisfaction - 2023 15th international congress on advanced applied informatics winter (iiai-aai-winter). pages 82–85. DOI: https://doi.org/10.1109/IIAI-AAI-Winter61682.2023.00023.

Rani, S., Mittal, A., and Singh, H. (2024). Enhancing user experience and satisfaction in robo-advisor platforms: A multi-dimensional analysis of usability factors and interface design elements - 2024 international conference on communication, computer sciences and engineering (ic3se). pages 1726–1730. DOI: https://doi.org/10.1109/IC3SE62002.2024.10593038.

Rapp, A., Curti, L., and 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. DOI: https://doi.org/10.1016/j.ijhcs.2021.102630.

Rocha, M., Soares, P., Caldeira, L., Souza, P., Vilela, R., Valle, P., and Silva, W. (2025). Tecnologias de avaliação da experiência do usuário (ux) em chatbots: Um mapeamento sistemático da literatura. In *Anais do II Workshop sobre Bots na Engenharia de Software*, pages 21–30, Porto Alegre, RS, Brasil. SBC. DOI: https://doi.org/10.5753/wbots.2025.15180.

Rodríguez-Ruiz, J., Marín-López, I., and Espejo-Siles, R. (2025). Is artificial intelligence use related to self-control, self-esteem and self-efficacy among university students? *Education and Information Technologies*, 30(2):2507–2524. DOI: https://doi.org/10.1007/s10639-024-12906-6.

Samonte, M. J. C. and Lising, E. C. (2024). Design analysis in improving customer relations using conversational artificial intelligence - 2024 14th international conference on software technology and engineering (icste). pages 144–148. DOI: https://doi.org/10.1109/ICSTE63875.2024.00032.

Saoud, L. S. and Romman, A. A. (2024). Ai-powered chatbots in the telecommunications industry in the uae: Studying the factors impacting consumer adoption - 2024 15th international conference on information and communication technology convergence (ictc). pages 1549–1553. DOI: https://doi.org/10.1109/ICTC62082.2024.10827670.

Sauka, K., Saou, Y., Situmeang, F. B. I., and Kackovic, M. (2026). A nexus of explainability and anthropomorphism in ai-chatbots. In *Explainable Artificial Intelligence*, pages 115–138. Springer Nature Switzerland. DOI: https://doi.org/10.1007/978-3-032-08317-3_6.

Shahid, A., Chung, J., and Heo, S. (2025). Exploring older adults personality preferences for llm-powered conversational companions. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. DOI: https://doi.org/10.1145/3706599.3719976.

Shan, Y., Ji, M., Xie, W., Zhang, X., Qian, X., Li, R., and Hao, T. (2022). Use of health care chatbots among young people in china during the omicron wave of covid-19: Evaluation of the user experience of and satisfaction with the technology. *JMIR Human Factors*, 9(2):e36831. DOI: https://doi.org/10.2196/36831.

Siddals, S., Torous, J., and Coxon, A. (2024). "It happened to be the perfect thing": experiences of generative AI chatbots for mental health. *npj Mental Health Research*, 3(1):48. DOI: https://doi.org/10.1038/s44184-024-00097-4.

Silva, G. and Canedo, E. (2024). Human factors in the design of chatbot interactions: Conversational design practices. In *Anais do XXIII Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*, pages 487–498, Porto Alegre, RS, Brasil. SBC. DOI: https://doi.org/10.1145/3702038.3702083.

Silva, G. R. S. and Canedo, E. D. (2025). Privacy in chatbot conversation-driven development: A comprehensive review and requirements proposal. *ACM Transactions on Software Engineering and Methodology*. DOI: https://doi.org/10.1145/3730578.

Skjuve, M., Følstad, A., Fostervold, K. I., and Brandtzaeg, P. (2021). My chatbot companion - a study of human-chatbot relationships. *International Journal of Human-Computer Studies*. DOI: https://doi.org/10.1016/j.ijhcs.2021.102601.

Thalhammer, P. T. (2025). Generative ai meets accessibility: Deformable interfaces and multimodal solutions. In *Proceedings of the Nineteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '25, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3689050.3704798.

Tsumura, T. and Yamada, S. (2025). Shaping empathy and trust toward agents: The role of agent behavior modification and attitude. *IEEE Access*, 13:116908–116923. DOI: https://doi.org/10.1109/ACCESS.2025.3584456.

Verma, A. K., Sarkar, T., Rakhra, M., Pandey, V. K., and Chakrabarty, P. (2024). Studying the impact of voice-activated chatbot's with artificial intelligence on consumer interaction in india - 2024 4th international conference on technological advancements in computational sciences (ictacs). pages 1708–1713. DOI: https://doi.org/10.1109/ICTACS62700.2024.10841074.

Virvou, M. and Tsihrintzis, G. A. (2024). Artificial intelligence chatbots and social robots in education: Feper framework for efficiency, pedagogical and ethical requirements - 2024 15th international conference on information, intelligence, systems applications (iisa). pages 1–8. DOI: https://doi.org/10.1109/IISA62523.2024.10786710.

Wang, H., Gupta, S., Singhal, A., Muttreja, P., Singh, S., Sharma, P., and Piterova, A. (2022). An artificial intelligence chatbot for young people's sexual and reproductive health in india (snehai): Instrumental case study. *Journal of Medical Internet Research*, 24(1):e29969. DOI: https://doi.org/10.2196/29969.

Wang, X., Wang, X., Park, S., and Yao, Y. (2025). Mental models of generative ai chatbot ecosystems. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*, IUI '25, page 1016–1031, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3708359.3712125.

Weber, P. and Ludwig, T. (2020). (non-)interacting with conversational agents: perceptions and motivations of using chatbots and voice assistants - proceedings of mensch und computer 2020. page 321–331. DOI: https://doi.org/10.1145/3404983.3405513.

Wicaksono, A., Barunastra, F. I., Suhari, M. C. J. G., and Jingga, F. (2024). Students level of trust in the use of ai chatbots in higher education: A quantitative analysis adopting technology acceptance model - 2024 7th international conference on contemporary computing and informatics (ic3i). 7:305–310. DOI: https://doi.org/10.1109/IC3I61595.2024.10829196.

Wikipedia contributors (2024). Chatbot — Wikipedia, The Free Encyclopedia. `https://en.wikipedia.org/w/index.php?title=Chatbot&oldid=1228514106`. Accessed: 23 August 2026.

Wikipedia contributors (2025). Artificial intelligence — Wikipedia, The Free Encyclopedia. `https://en.wikipedia.org/wiki/Artificial_intelligence`. Accessed: 23 August 2026.

Wong, J., Foussat, A. C., Ting, S., Acerbi, E., van Elburg, R. M., and Chua, M. C. (2021). A chatbot to engage parents of preterm and term infants on parental stress, parental sleep, and infant feeding: Usability and feasibility

study. *JMIR Pediatrics and Parenting*, 4(4):e30169. DOI: https://doi.org/10.2196/30169.

Yeh, P.-L., Kuo, W.-C., Tseng, B.-L., and Sung, Y.-H. (2025). Does the AI-driven Chatbot Work? Effectiveness of the Woebot app in reducing anxiety and depression in group counseling courses and student acceptance of technological aids. *Current Psychology*. DOI: https://doi.org/10.1007/s12144-025-07359-0.

Zhao, X. (2025). Factors influencing medical students' adoption of ai educational agents: an extended utaut model. *BMC Medical Education*, 25(1):article 8234. DOI: https://doi.org/10.1186/s12909-025-08234-z.

Zimmerman, A., Janhonen, J., and Beer, E. (2024). Human/AI relationships: challenges, downsides, and impacts on human/human relationships. *AI and Ethics*, 4(4):1555–1567. DOI: https://doi.org/10.1007/s43681-023-00348-8.

Table 9. Summary of the reviewed articles (S1 to S14)

Id (Citation Key)	Title	Publication Location	Full Authors	DOI / URL
[S1] [Fabio *et al*., 2024]	AI-based chatbot interactions and critical thinking skills: an exploratory study	Current Psychology	Rosa Angela Fabio, Alessio Plebe, Rossella Suriano	https://doi.org/10.1007/s12144-024-06795-8
[S2] [May and Denecke, 2021]	AI Bots: Navigating the Next Wave of Innovation	Proc. ICRITO 2024	MD Samir Akhtar, Gauri Mathur, Manik Rakhra, Chirag Sharma	https://doi.org/10.1109/ICEECT61758.2024.10738911
[S3] [Khurana *et al*., 2021]	ChatrEx: Designing Explainable Chatbot Interfaces for Enhancing Usefulness, Transparency, and Trust	Proc. VL/HCC 2021	Anjali Khurana, Parsa Alamzadeh, Parmit K. Chilana	https://doi.org/10.1109/VL/HCC51201.2021.9576440
[S4] [Joshi and Bengler, 2024]	Crafting Human-AI Interaction: A Rhetorical Approach to Adaptive Interaction in Conversational Agents	Proc. ACM HAI 2024	Rutuja Joshi, Klaus Bengler	https://doi.org/10.1145/3687272.3688297
[S5] [Meng *et al*., 2024]	Developing Design Features to Facilitate AI-Assisted User Interactions	Proc. IEEE SIEDS 2024	S. Meng, R. Dollahite, A. Kaur, P. Schell, A. Sharma, G. Ventre	https://doi.org/10.1109/SIEDS61124.2024.10534665
[S6] [Mubassira *et al*., 2025]	Enhancing EmoBot: An In-Depth Analysis of User Expectation and Satisfaction in an Emotion-Aware Chatbot	Proc. NSysS 2024	Taseen Mubassira, Mehedi Hasan, Jannatun Noor, A. B. M. Alim Al Islam	https://doi.org/10.1145/3704522.3704527
[S7] [Zimmerman *et al*., 2024]	Human/AI relationships: challenges, downsides, and impacts on human/human relationships	AI and Ethics	Anne Zimmerman, Joel Janhonen, Emily Beer	https://doi.org/10.1007/s43681-023-00348-8
[S8] [Weber and Ludwig, 2020]	((Non-)Interacting with conversational agents: perceptions and motivations of using chatbots and voice assistants	Mensch und Computer 2020	Philip Weber, Thomas Ludwig	https://doi.org/10.1145/3404983.3405513
[S9] [Min *et al*., 2021]	Research on Users' Trust of Chatbots Driven by AI: An Empirical Analysis Based on System Factors and User Characteristics	Proc. ICCECE 2021	Fan Min, Yucan He, Zou Fang, Jiang Xuan	https://doi.org/10.1109/ICCECE51280.2021.9342098
[S10] [Aumüller *et al*., 2024]	Rethinking feminized service bots: user responses to abstract and gender-ambiguous chatbot avatars in a large-scale interaction study	Personal and Ubiquitous Computing	Anna Aumüller, Andreas Winklbauer, Beatrice Schreibmaier, Bernad Batinic & Martina Mara	https://doi.org/10.1007/s00779-024-01830-8
[S11] [Kim *et al*., 2023]	AI increases unethical consumer behavior due to reduced anticipatory guilt	J. Academy Mark. Science	TaeWoo Kim, Hyejin Lee, Michelle Yoosun Kim, SunAh Kim & Adam Duhachek	https://doi.org/10.1007/s11747-021-00832-9
[S12] [Saoud and Romman, 2024]	AI-Powered Chatbots in the Telecommunications Industry in the UAE: Studying the Factors Impacting Consumer Adoption	Proc. ICTC 2024	Loubna Saad Saoud, Ali Abu Romman	https://doi.org/10.1109/ICTC62082.2024.10827670
[S13] [Putra Antonio *et al*., 2023]	Analyzing the Impact of Customer Service Chatbots on User Satisfaction	Proc. IIAI-AAI 2023	Elkana Rahadian Putra Antonio, Muhammad Farras Fadhilah, Fikri Faiq; Renaldy Fredyan,Hady Pranoto	https://doi.org/10.1109/IIAI-AAI-Winter61682.2023.00023
[S14] [Balaji and Rao, 2024]	AI and Chatbots: Transforming User Experience in E-tailing	Proc. ICCES 2024	K. Balaji, P. Seshagiri Rao	https://doi.org/10.1109/ICCES63552.2024.10859474

Table 10. Summary of the reviewed articles (S15 to S28)

Id (Citation Key)	Title	Publication Location	Full Authors	DOI / URL
[S15] [Rani *et al*., 2024]	Enhancing User Experience and Satisfaction in Robo-Advisor Platforms: A Multi-Dimensional Analysis of Usability Factors and Interface Design Elements	Proc. IC3SE 2024	Sonam Rani, Ajit Mittal, Harjit Singh	https://doi.org/10.1109/IC3SE62002.2024.10593038
[S16] [Ding and Najaf, 2024]	Interactivity, humanness, and trust: a psychological approach to AI chatbot adoption in e-commerce	BMC Psychology	Yi Ding, Muzammil Najaf	https://doi.org/10.1186/s40359-024-02083-z
[S17] [Verma *et al*., 2024]	Studying the Impact of Voice-Activated Chatbot's with Artificial Intelligence on Consumer Interaction in India	Proc. ICTACS 2024	Ashish Kumar Verma, Tiyas Sarkar, Manik Rakhra, Vikas Kumar Pandey, Projjal Chakrabarty	https://doi.org/10.1109/ICTACS62700.2024.10841074
[S18] [Liu *et al*., 2022]	The Effect of Chatbot-customer Interaction on Consumer Brand Advocacy: Exploring the Role of Chatbots	Proc. ICEIEC 2022	Yang Liu, Xiaolong Li, Zheng Xiang	https://doi.org/10.1109/ICEIEC54567.2022.9835050
[S19] [Rodríguez-Ruiz *et al*., 2025]	Is artificial intelligence use related to self-control, self-esteem and self-efficacy among university students?	Ed. and Inf. Tech.	Joaquín Rodríguez-Ruiz, Inmaculada Marín-López & Raquel Espejo-Siles	https://doi.org/10.1007/s10639-024-12906-6
[S20] [Virvou and Tsihrintzis, 2024]	Artificial Intelligence Chatbots and Social Robots in Education: FEPER Framework for Efficiency, Pedagogical and Ethical Requirements	Proc. IISA 2024	Maria Virvou, George A. Tsihrintzis	https://doi.org/10.1109/IISA62523.2024.10786710
[S21] [Jain *et al*., 2024]	Enhancing Educational Interactions: A Comprehensive Review of AI Chatbots in Learning Environments	Proc. ICRITO 2024	Vipin Jain, Isha Singh, Madiha Syed; ,Sweety Mondal, Deepanshu Ranjan Palai	https://doi.org/10.1109/ICRITO61523.2024.10522392
[S22] [Kothari and Noel Newton Fernando, 2024]	Enhancing Human-Computer Interaction through AI: A Study on ChatGPT in Educational Environments	Proc. IEEE CAI 2024	Dhurval Kothari, Owen Fernando	https://doi.org/10.1109/CAI59869.2024.00100
[S23] [Karataş *et al*., 2024]	Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners	Education and Information Technologies	Fatih Karataş, Faramarz Yaşar Abedi, Filiz Ozek Gunyel, Derya Karadeniz, Yasemin Kuzgun	https://doi.org/10.1007/s10639-024-12574-6
[S24] [Wicaksono *et al*., 2024]	Students Level of Trust in the Use of AI Chatbots in Higher Education: A Quantitative Analysis Adopting Technology Acceptance Model	Proc. IC3I 2024	Arief Wicaksono, Farrel Izza Barunastra, Michael Chandra Jaya, Giovani Suhari, Fredy Jingga	https://doi.org/10.1109/IC3I61595.2024.10829196
[S25] [Yeh *et al*., 2025]	Does the AI-driven Chatbot Work? Effectiveness of the Woebot app in reducing anxiety and depression in group counseling courses and student acceptance of technological aids	Current Psychology	Pao-Ling Yeh, Wen-Cheng Kuo, Bei-Lu Tseng, Yu-Hsien Sung	https://doi.org/10.1007/s12144-025-07359-0
[S26] [Peng *et al*., 2024]	How loneliness leads to the conversational AI usage intention: the roles of anthropomorphic interface, para-social interaction	Current Psychology	Chen Peng, Shengyu Zhang, Fang Wen, Kun Liu	https://doi.org/10.1007/s12144-024-06809-5
[S27] [Park *et al*., 2024]	Human vs. machine-like representation in chatbot mental health counseling: the serial mediation of psychological distance and trust on compliance intention	Current Psychology	Gain Park, Jiyun Chung, Seyoung Lee	https://doi.org/10.1007/s12144-023-04653-7
[S28] [Siddals *et al*., 2024]	"It happened to be the perfect thing": experiences of generative AI chatbots for mental health	npj Mental Health Res.	Steven Siddals, John Torous, Astrid Coxon	https://doi.org/10.1038/s44184-024-00097-4

Table 11. Summary of the reviewed articles (S29 to S42)

Id (Citation Key)	Title	Publication Location	Full Authors	DOI / URL
[S29] [Pozzi and De Proost, 2024]	Keeping an AI on the mental health of vulnerable populations: reflections on the potential for participatory injustice	AI and Ethics	Giorgia Pozzi, Michiel De Proost	https://doi.org/10.1007/s43681-024-00523-5
[S30] [Maples *et al*., 2024]	Loneliness and suicide mitigation for students using GPT3-enabled chatbots	npj Mental Health Research	Bethanie Maples, Merve Cerit, Aditya Vishwanath & Roy Pea	https://doi.org/10.1038/s44184-023-00047-6
[S31] [Haugeland *et al*., 2022]	Understanding the user experience of customer service chatbots: An experimental study of chatbot interaction design	International Journal of Human-Computer Studies	Isabel Kathleen Fornell Haugeland, Asbjørn Følstad, Cameron Taylor, Jan Alexander Steinholt Nordheim	https://doi.org/10.1016/j.ijhcs.2022.102788
[S32] [Kang and Hong, 2025]	Development and Evaluation of a Mental Health Chatbot Using ChatGPT 4.0: Mixed Methods User Experience Study With Korean Users	JMIR Medical Informatics	Boyoung Kang, Munpyo Hong	https://doi.org/10.2196/63538
[S33] [Shan *et al*., 2022]	Use of Health Care Chatbots Among Young People in China During the Omicron Wave of COVID-19: Evaluation of the User Experience of and Satisfaction With the Technology	JMIR Human Factors	Yi Shan, Meng Ji, Wenxiu Xie, Xiaomin Zhang, Xiaobo Qian, Rongying Li, Tianyong Hao	https://doi.org/10.2196/36831
[S34] [He *et al*., 2022]	Mental Health Chatbot for Young Adults With Depressive Symptoms During the COVID-19 Pandemic: Single-Blind, Three-Arm Randomized Controlled Trial	Journal of Medical Internet Research (JMIR)	Yuhao He, Li Yang, Xiaokun Zhu, Bin Wu, Shuo Zhang, Chunlian Qian, Tian Tian	https://doi.org/10.2196/40719
[S35] [Wang *et al*., 2022]	An Artificial Intelligence Chatbot for Young People's Sexual and Reproductive Health in India (SnehAI): Instrumental Case Study	Journal of Medical Internet Research (JMIR)	Hua Wang, Sneha Gupta, Arvind Singhal, Poonam Muttreja, Sanghamitra Singh, Poorva Sharma, Alice Piterova	https://doi.org/10.2196/29969
[S36] [Chung *et al*., 2021]	A Chatbot for Perinatal Women's and Partners' Obstetric and Mental Health Care: Development and Usability Evaluation Study	JMIR Medical Informatics	Kyungmi Chung, Hee Young Cho, Jin Young Park	https://doi.org/10.2196/18607
[S37] [Ludin *et al*., 2022]	A Chatbot to Support Young People During the COVID-19 Pandemic in New Zealand: Evaluation of the Real-World Rollout of an Open Trial	Journal of Medical Internet Research (JMIR)	Nicola Ludin, Chester Holt-Quick, Sarah Hopkins, Karolina Stasiak, Sarah Hetrick, Jim Warren, Tania Cargo	https://doi.org/10.2196/38743
[S38] [Wong *et al*., 2021]	A Chatbot to Engage Parents of Preterm and Term Infants on Parental Stress, Parental Sleep, and Infant Feeding: Usability and Feasibility Study	JMIR Pediatrics and Parenting	Jill Wong, Agathe C. Foussat, Steven Ting, Enzo Acerbi, Ruurd M. van Elburg, Chua Mei Chien	https://doi.org/10.2196/30169
[S39] [Chagas *et al*., 2023]	Evaluating User Experience With a Chatbot Designed as a Public Health Response to the COVID-19 Pandemic in Brazil: Mixed Methods Study	JMIR Human Factors	Bruno Azevedo Chagas, Adriana Silvina Pagano, Raquel Oliveira Prates, Elisa Cordeiro Praes, Kícila Ferreguetti, Helena Vaz, Zilma Silveira Nogueira Reis, Leonardo Bonisson Ribeiro, Antonio Luiz Pinho Ribeiro, Thais Marques Pedroso, Alline Beleigoli, Clara Rodrigues Alves Oliveira, Milena Soriano Marcolino.	https://doi.org/10.2196/43135
[S40] [Adnin *et al*., 2025]	Examining Student and Teacher Perspectives on Undisclosed Use of Generative AI in Academic Work	Proc. ACM CHI 2025	Adnin, Rudaiba and Pandkar, Atharva and Yao, Bingsheng and Wang, Dakuo and Das, Maitraye	https://doi.org/10.1145/3706598.3713393
[S41] [Sauka *et al*., 2026]	A Nexus of Explainability and Anthropomorphism in AI-Chatbots	Springer Nature Switzerland	Sauka, Kudzai and Saou, Youssef and Situmeang, Frederik B. I. and Kackovic, Monika	https://doi.org/10.1007/978-3-032-08317-3_6
[S42] [Genç *et al*., 2025]	Persuasion in Pixels and Prose: The Effects of Emotional Language and Visuals in Agent Conversations on Decision-Making	Proc. ACM CHI 2025	Genç, Hüseyin Uğur and Chandrasegaran, Senthil and Dingler, Tilman and Verma, Himanshu	https://doi.org/10.1145/3706598.3713579

Table 12. Summary of the reviewed articles (S43 to S56)

Id (Citation Key)	Title	Publication Location	Full Authors	DOI / URL
[S43] [Shahid *et al*., 2025]	Exploring Older Adults Personality Preferences for LLM-powered Conversational Companions	Proc. ACM CHI EA 2025	Shahid, Ajwa and Chung, Jane and Heo, Seongkook	https://doi.org/10.1145/3706599.3719976
[S44] [Zhao, 2025]	Factors influencing medical students' adoption of AI educational agents: an extended UTAUT model	BMC Medical Education	Xiaoxiong Zhao, Wankun Liu, Sansi Yue, Jia Chen, Dan Xia, Kaijian Bing, Xiaotian Xia and Keshan Wang	https://doi.org/10.1186/s12909-025-08234-z
[S45] [Wang *et al*., 2025]	Mental Models of Generative AI Chatbot Ecosystems	Proc. ACM IUI 2025	Wang, Xingyi and Wang, Xiaozheng and Park, Sunyup and Yao, Yaxing	https://doi.org/10.1145/3708359.3712125
[S46] [He *et al*., 2025]	Is Conversational XAI All You Need? Human-AI Decision Making With a Conversational XAI Assistant	ACM IUI 2025	He, Gaole and Aishwarya, Nilay and Gadiraju, Ujwal	https://doi.org/10.1145/3708359.3712133
[S47] [Marmolejo-Ramos *et al*., 2026]	University students' perceptions and adoptions of AI: a cross-national study	Discover Artificial Intelligence	Marmolejo-Ramos, Fernando and Abadia, Rhoda and Karakale, Özge and Barrera-Causil, Carlos and Männikkö, Niko and Ghorbani, Babak Daneshvar and Strzelecki, Artur and Nwizu, Stella and Tavares, Célia and Castillo, Mauricio and Escajal, Bruno and Rodriguez, Pilar and Som, Bidisha and Ngo, Van Giang and Özdoğru, Asil Ali and Ariyabuddhiphongs, Kris and Thongrin, Santat and Tejada, Julian	https://doi.org/10.1007/s44163-026-01042-4
[S48] [Atcheson *et al*., 2025]	"I'd Never Actually Realized How Big An Impact It Had Until Now": Perspectives of University Students with Disabilities on Generative Artificial Intelligence	Proc. ACM CHI 2025	Atcheson, Alex and Khan, Omar and Siemann, Brian and Jain, Anika and Karahalios, Karrie	https://doi.org/10.1145/3706598.3714121
[S49] [Kang *et al*., 2025]	ExplainitAI: When do we trust artificial intelligence? The influence of content and explainability in a cross-cultural comparison	Proc. ACM CHI EA 2025	Kang, Sora and Potinteu, Andreea-Elena and Said, Nadia	https://doi.org/10.1145/3706599.3720222
[S50] [Mayerhofer *et al*., 2025]	Blending Queries and Conversations: Understanding Trust, Verification, and System Choice in Search and Chat Interactions	Proc. ACM CHIIR 2025	Mayerhofer, Kerstin and Capra, Rob and Elsweiler, David	https://doi.org/10.1145/3698204.3716454
[S51] [Alshammari *et al*., 2025]	Determinants of student adoption of artificial intelligence applications in higher education	Scientific Reports	Alshammari, Sultan Hammad and Alrashidi, Muna Eid and Alshammari, Mohammed Habib and Alshammari, Ali Essa A. and Alkhwaldi, Abeer F.	https://doi.org/10.1038/s41598-025-19851-5
[S52] [Alshammari and Babu, 2025]	The mediating role of satisfaction in the relationship between perceived usefulness, perceived ease of use and students' behavioural intention to use ChatGPT	Scientific Reports	Alshammari, Sultan Hammad and Babu, Eldho	https://doi.org/10.1038/s41598-025-91634-4
[S53] [Jung *et al*., 2025]	'I've talked to ChatGPT about my issues last night.': Examining Mental Health Conversations with Large Language Models through Reddit Analysis	Proceedings of the ACM on Human-Computer Interaction Vol. 9, No. 7	Jung, Kyuha and Lee, Gyuho and Huang, Yuanhui and Chen, Yunan	https://doi.org/10.1145/3757537
[S54] [Hasselwander and Lah, 2026]	Agentic AI Arrives: How Gen Z Adopts Autonomous AI Agents	IEEE Access	Hasselwander, Marc and Lah, Oliver	https://doi.org/10.1109/ACCESS.2026.3665348
[S55] [Tsumura and Yamada, 2025]	Shaping Empathy and Trust Toward Agents: The Role of Agent Behavior Modification and Attitude	IEEE Access	Tsumura, Takahiro and Yamada, Seiji	https://doi.org/10.1109/ACCESS.2025.3584456
[S56] [Thalhammer, 2025]	Generative AI Meets Accessibility: Deformable Interfaces and Multimodal Solutions	Proc. ACM TEI 2025	Thalhammer, Philipp Tim	https://doi.org/10.1145/3689050.3704798