Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 4 Issue 3 (2024)

A Systematic Analysis on Chatbot Information Security

Shivani Vats[1*], Deepika[2]

[1,2]Assistant Professor, Jagan Institute of Management Studies (JIMS), Sector- 5, Rohini,
[*]Email: drshivanivats@gmail.com, Email: deepikagahlan@gmail.com, [*](https://orcid.org/0009-0002-7174-4165),
(https://orcid.org/0000-0003-2560-5690*)*

***Corresponding author:** Shivani Vats
*Assistant Professor, Jagan Institute of Management Studies (JIMS), Sector- 5, Rohini,
[*]Email: drshivanivats@gmail.com

In recent years, chatbots have grown in popularity, but they also pose security threats and weaknesses that require attention. This systematic literature analysis reviews the prior work on information security in chatbots, highlighting potential vulnerabilities, suggested countermeasures, and prospective future research approaches. According to the analysis, there are a number of security risks that chatbots must guard against, including malicious input, user profiling, contextual attacks, and data breaches. However, these problems may be addressed with the help of organizational controls, end-to-end encryption, and blockchain technology. The evaluation also emphasizes how crucial it is to uphold user confidence and allay privacy worries in order for chatbots to be adopted and used successfully in the future. This review's taxonomy offers a helpful framework for classifying the papers and their findings.

Keywords: Chatbot; information security; systematic literature review (SLR); ChatGPT; security

1. Introduction

Conversational agents are another name for chatbots[1]. They are software applications that use artificial intelligence, natural language processing, and machine learning to mimic human speech[2]. A wide range of businesses and applications, including e-commerce, healthcare, banking, and education, have benefited from the emergence of chatbots by becoming more convenient and effective. However, as these systems spread more widely, they also become more open to various security risks and intrusions, which raises questions about the security and privacy of critical user data[3].

Information security has actually drawn more attention recently [4, 5]. The safeguarding of users' private information is one of the main information security concerns faced by chatbots [6, 7,8,9,10,11]. The quantity of personal information exchanged by chatbots, including financial data, health information, and personally identifiable information, is growing as they are utilized more often across a variety of sectors and applications. This makes them a desirable target for fraudsters, who might try to utilize chatbot weaknesses to get access to customer data without authorization.

For instance, if a healthcare chatbot is breached, an attacker might have access to private patient data like medical records, prescriptions, and other details. Similar to the previous example, if a finance chatbot is compromised [**12**,**13**], an attacker may obtain user financial information, including credit card numbers, bank account information, and transaction history. The requirement to uphold user confidence and trust in these systems is a crucial component of chatbot information security[**6**,**7**,**8**,**9**,**14**,**15**,**16**,**17**,**18**]. When using chatbots, users must have faith that their personal information is safe and secured. Trust among users can be damaged by a security lapse or data leak, which can have serious repercussions for companies and organizations who employ chatbots to offer customer service and assistance.

For instance, the businesses impose secret regulations in the terms and conditions and disclaimers of the website to obtain user agreement to use their data. Companies in this situation are legally permitted to store personal data. However, it's possible that the user is unaware that their data has been shared with outside parties [19, 20]. The fact that older adults appear to behave differently than younger adults when selecting a chatbot for customer assistance as opposed to connecting to a live agent is another factor to take into particular account for consumer trust [21,22]. Although they make up a sizable portion of the population, older folks may still value human interaction more than chatbots, which means that relying primarily or mostly on chatbot communication risks alienating this market [23].

The nature of security threats and vulnerabilities is constantly changing, which presents significant research problems [6,7,8,9,10,16]. New attack types that take advantage of previously undiscovered weaknesses might appear as chatbots get more sophisticated and capable. To recognize and reduce emerging dangers, this necessitates continuous study and development. Supercomputers from Microsoft Azure AI were used to train GPT-4, which is superior to ChatGPT. It employs a deep learning methodology to build ever-more complex and effective language models by utilizing more data and more computation [24]. To predict results, it must first be trained on sizable datasets of data from the relevant domains (such as patient and user conversation data) [25]. A chatbot cannot, however, learn from encrypted data. As a result, when decrypting data for training, there is a chance that it will be disclosed to outside parties.

Furthermore, developing secure chatbots necessitates a multidisciplinary approach that takes into account not just technological security measures but also user trust, privacy, and ethical considerations [6,7,8,9,14,15,16,17]. Developers must address the possible social impact of chatbots and ensure that they are created and deployed ethically and responsibly [26]. This can be a difficult process that necessitates the cooperation of security professionals, developers, policymakers, and users.

Moreover, the diverse contexts and situations of chatbots present particular security challenges which need to be solved in specific ways. A chatbot used in healthcare, for example, may require different security precautions than one used in banking or e-commerce [27,28,29]. In this setting, a comprehensive approach to security is required, taking into account the unique needs and requirements of each use case. A chatbot used in healthcare, for example, may require different security precautions than one used in banking or e-commerce [27,28,29]. In this setting, a comprehensive approach to security is required, taking into account the unique needs and requirements of each use case. The purpose of this research was to present a complete analysis of the primary security threats and vulnerabilities encountered by chatbots, as well as to highlight the techniques and technologies that can be utilized to minimize these risks. Chatbots in e-commerce aim to enhance customer service, but their effectiveness depends on using human-like design cues to increase user compliance [37]. AI chatbots in education can enhance personalized learning by providing instant feedback and analyzing learner data, leading to improved student satisfaction and performance, though they also raise ethical and privacy concerns [38].

2. Method

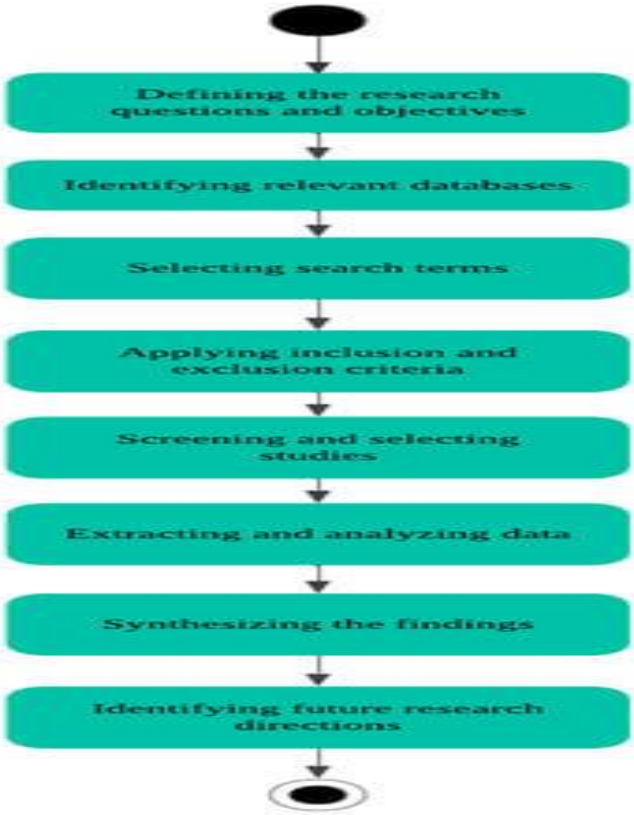

Defining the research questions and objectives
Identifying relevant databases
Selecting search terms
Applying inclusion and exclusion criteria
Screening and selecting studies
Extracting and analyzing data
Synthesizing the findings
Identifying future research directions

Figure 1. Literature review methodology

Figure 1 depicts the methods utilized in this study for conducting a literature review. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines in conducting this systematic literature review on chatbots and security to guarantee a full and transparent examination of the material.

PRISMA is a widely established and acknowledged methodology for performing systematic literature reviews (SLRs) [30]. It offers a clear and repeatable structure for conducting literature searches, filtering and selecting relevant articles, and synthesizing the results.

The PRISMA technique was regarded an appropriate approach in the context of an SLR focused on security threats and vulnerabilities in chatbots since it helps to ensure a complete and systematic search of the literature as well as a rigorous process for evaluating and choosing relevant articles. Furthermore, the PRISMA approach includes a detailed reporting checklist that can assist in ensuring that the review is presented in a clear and transparent manner.

Other techniques to SLRs, such as the Cochrane methodology [31], may be useful for certain study issues or topics. However, given the emphasis on transparency and reproducibility, we believe that the PRISMA technique was the best fit for this unique SLR.

The literature review process consists of eight consecutive steps, which are as follows:

Step 1. Defining the research questions and objectives: To guide the search and study of the literature, the research questions and objectives were defined. The following are the research questions and objectives:

Research Questions:

1. What are the most serious security dangers and weaknesses that chatbots face?
2. What techniques and technology are available to help manage these risks?

Objectives:

1. To provide a complete overview of the primary security threats and weaknesses that chatbots confront.
2. To emphasize the tactics and technologies available for mitigating these hazards.

The findings of this literature review, we feel, can influence future study and practice in the topic of information security in chatbots.

Step 2: **Identifying relevant databases:** To ensure a thorough search of the literature, several databases were used, including the ACM Digital Library, IEEE Xplore, ScienceDirect, and Web of Science.

Step 3: Selecting search terms: The research questions and objectives guided the selection of search phrases. "chatbot" OR "ChatGPT" AND "security" OR "information security" were the search phrases used.

Step 4: Applying inclusion and exclusion criteria:

Inclusion Criteria: We included peer-reviewed research articles published between 2016 and 2023 that focused particularly on the security of chatbots. Furthermore, we only examined papers that were published in English and were available in full-text format.

Exclusion Criteria: We omitted studies that did not focus on chatbot security, such as those that focused on chatbot development, technology, and applications, natural language processing, or user experience. This review assessed only peer-reviewed research articles for inclusion.

Step 5: Screening and selecting studies: The studies were chosen based on their relevance to the study questions and objectives. Following that, full-text papers were screened, and studies that did not fulfill the inclusion criteria or were irrelevant to the study topics were excluded.

Step 6: Extracting and analyzing data: Relevant data from the selected studies, such as the research techniques utilized, the types of chatbots investigated, and the specific security vulnerabilities addressed, were extracted. The gathered data was then evaluated for common themes, patterns, and gaps in the literature.

Step 7: Synthesizing the findings: The findings were synthesized and presented in a narrative fashion, with tables and figures providing visual representations of the data.

Step 8: Identifying future research directions: Future research directions were identified and provided in the paper's conclusion section based on the examination of the literature.

3. Result

The PRISMA flow diagram [30] for a systematic examination of chatbot security is shown in Figure 2. There are three processes involved (identification, screening, and eligibility). During the initial identification phase, we retrieved 1193 items via an electronic database search. After deleting duplicates, 179 articles remained. The screening process comprised reviewing the titles and abstracts of the publications, which resulted in the elimination of 62 articles. During the eligibility phase, full-text screening was performed on the remaining 19 articles, and 10 articles were rejected for failing to meet the inclusion criteria.

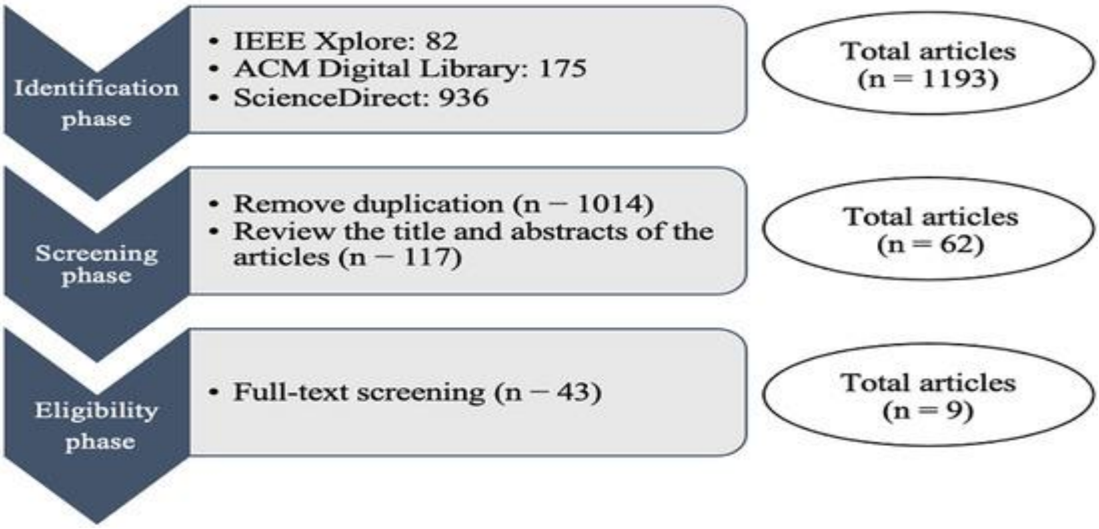

Identification phase
• IEEE Xplore: 82
• ACM Digital Library: 175
• ScienceDirect: 936
Total articles (n = 1193)
Screening phase
• Remove duplication (n – 1014)
• Review the title and abstracts of the articles (n – 117)
Total articles (n = 62)
Eligibility phase
• Full-text screening (n – 43)
Total articles (n = 9)

Figure 2. PRISMA flow diagram [30] for a systematic examination of chatbot security.

The nine publications that met the inclusion requirements were evaluated and synthesized to offer a high-level overview of the primary security concerns and vulnerabilities connected with chatbots, as well as the ways proposed to mitigate them. The findings of our systematic literature review are presented below.

Figure 3 depicts the papers and fields related to information security. We discovered various publications relating to security threats and vulnerabilities in chatbots through a rigorous literature review. Five articles in particular [6,7,8,10,16] were discovered to be connected to security issues and vulnerabilities in chatbots. These articles examined potential chatbot security threats such as malicious input, user profiling, contextual assaults, and data breaches. The articles stressed the necessity of developers being aware of these potential vulnerabilities and taking precautions to protect their systems against assaults.

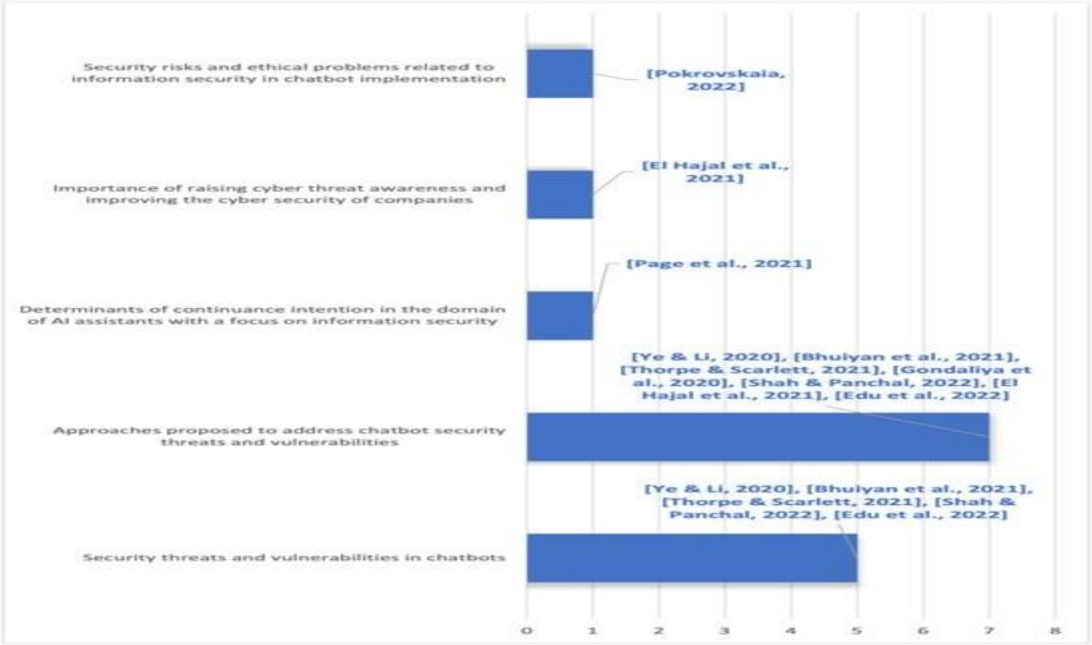


Figure 3: Information security articles and their respective fields [6,7,8,9,10,14,15,16,30]

The analysis also included seven articles [6,7,8,9,10,14,16] that discussed techniques to addressing chatbot security threats and vulnerabilities. These articles presented several methods to address chatbot security concerns, such as the use of blockchain technology, end-to-end encryption, and the inclusion of organizational, administrative, and technical controls in the service level agreement.

The review found one paper [17] on factors of persistence intention in the domain of AI assistants with an emphasis on information security. In order to secure the successful adoption and continuous usage of AI assistants, this article underlined the need of preserving user trust and addressing privacy concerns.

Furthermore, one paper [14] emphasized the significance of increasing cyber threat awareness and strengthening corporate cyber security. It presented an AI-based conversational bot that functions as a personal assistant to improve cyber threat awareness and give the most up-to-date information and training to a company's personnel.

The article [15] explored the security dangers and ethical issues associated with information security in chatbot implementation. It noted that human conduct gives instances of both appropriate and inappropriate behavior for neural networks, and it offered a mechanism for detecting socio-cultural and information security concerns by monitoring the interests and motives of users, specialists, and programmers.

The findings and related articles are shown in Table 1. Article [6] focuses on the chatbot's security and privacy flaws. The article discusses various flaws that could jeopardize the system's security and the user's privacy. Malicious input, user profiling, contextual attacks, and data breaches are examples. Attackers could enter harmful input into the chatbot, such as SQL injection or cross-site scripting, in order to attack system vulnerabilities. Chatbots may collect sensitive information about users, such as personal data or behavioral patterns, which attackers may use for identity theft or targeted assaults. Attackers could leverage the context of the discussion to manipulate the user or the chatbot, such as impersonating the chatbot or tricking the user into disclosing sensitive information. Chatbots may be subject to data breaches in which user or chatbot data is disclosed to unauthorized parties. According to the article, developers should be aware of these potential vulnerabilities and take precautions to protect their systems from assaults. Despite chatbots' immense promise for many applications, it is critical to solve these weaknesses to maintain user security and privacy.

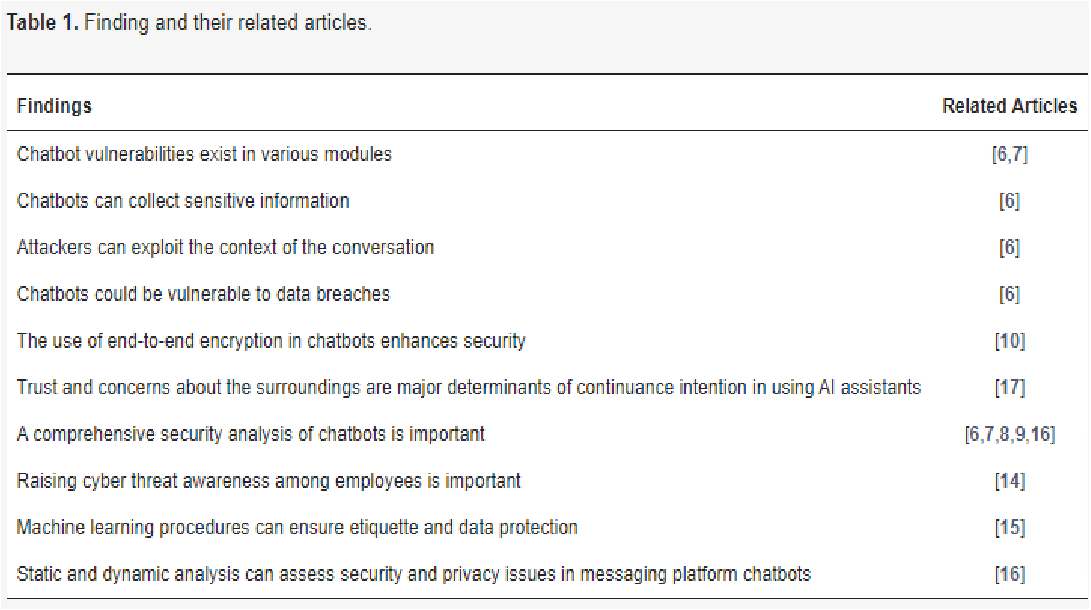


Table 1. Finding and their related articles.

Findings	Related Articles
Chatbot vulnerabilities exist in various modules	[6,7]
Chatbots can collect sensitive information	[6]
Attackers can exploit the context of the conversation	[6]
Chatbots could be vulnerable to data breaches	[6]
The use of end-to-end encryption in chatbots enhances security	[10]
Trust and concerns about the surroundings are major determinants of continuance intention in using AI assistants	[17]
A comprehensive security analysis of chatbots is important	[6,7,8,9,16]
Raising cyber threat awareness among employees is important	[14]
Machine learning procedures can ensure etiquette and data protection	[15]
Static and dynamic analysis can assess security and privacy issues in messaging platform chatbots	[16]

Some of the weaknesses discovered in chatbots are insecure authentication, data integrity difficulties, system availability, transparency, and privacy concerns, according to article [7]. Attackers may exploit these flaws to compromise a chatbot's security, obtain unauthorized access to sensitive information, or disrupt the chatbot's activities. The report suggests that blockchain technology be used to address some of these security vulnerabilities in financial chatbots. However, depending on the exact implementation and environment, chatbot vulnerabilities may vary, and other studies may uncover new or different forms of vulnerabilities.

The risks in chatbots are not covered directly in article [8]. The paper offers a concept and structure for a cyber-aware chatbot service that identifies and assists in the prevention of malware behavior on the user's PC. The malevolent behavior represented in the study is based on the vulnerabilities identified in the top ten of the Open Web Application Security

Project (OWASP), which details the Web's security issues. However, the paper makes no mention of chatbot vulnerabilities.

Article [9] discusses the potential dangers involved with chatbots, such as the confidentiality and integrity of user data, the dependability of chatbot responses, and the service level agreement (SLA) established by chatbot providers. The proposed checklist gives security administrators a mechanism to identify these risks prior to implementing chatbots. The paper also recommends a set of controls, such as organizational, managerial, and technological controls, that can be included in the SLA to manage these risks, and provides examples of how these controls can address specific risk factors, such as DDoS assaults on third-party infrastructure. Before signing a SLA with a chatbot provider, clients can be informed about the risks connected with the service using the proposed analysis.

Article [10] discusses chatbot security vulnerabilities, which could allow thieves or hackers to access information passing through the chatbot interface. As a result, the authors offer a method to secure chatbots by combining authentication (session) timeouts with encryption mechanisms such as the Double Ratchet algorithm modified with Paillier Cryptosystems. The suggested chatbot's use of end-to-end encryption ensures that only the intended receivers may decrypt messages, protecting an organization's critical data. A safe educational chatbot's major purpose is to protect kids' data from cyber-criminals, hackers, or attackers, so the interaction is completely secure utilizing end-to-end encryption (E2EE), protecting data privacy, confidentiality, integrity, and authentication.

The necessity of increasing cyber threat awareness and strengthening corporate cyber security by focusing on the weakest link—the human element layer—is discussed in article [14]. The study suggests an AI-based conversational bot that functions as a personal assistant to increase cyber threat awareness and provide the most up-to-date information and training to a company's personnel. The bot is intended to communicate with the user via WhatsApp and is capable of keeping track of each employee, rating their progress, and recommending training to improve shortcomings. The bot's implementation has had a significant impact on the staff, and the bot may update its database of any security breach and recommend methods to behave in the event of an attack. The necessity of increasing cyber threat awareness and strengthening corporate cyber security by focusing on the weakest link—the human element layer—is discussed in article [14]. The study suggests an AI-based conversational bot that functions as a personal assistant to increase cyber threat awareness and provide the most up-to-date information and training to a company's personnel. The bot is intended to communicate with the user via WhatsApp and is capable of keeping track of each employee, rating their progress, and recommending training to improve shortcomings. The bot's implementation has had a significant impact on the staff, and the bot may update its database of any security breach and recommend methods to behave in the event of an attack.

The article [15] explores the security dangers and ethical issues associated with information security in chatbot implementation. It stresses how human behavior provides neural networks with instances of both appropriate and incorrect conduct, and it suggests a mechanism for detecting socio-cultural and information security threats by monitoring the interests and motives of users, specialists, and programmers. The essay discusses critical techniques to maintaining etiquette and data safety and proposes machine learning procedures for chatbots in business ecosystems. The essay proposes that society members devise a system of cultural transmission, knowledge transfer, and civic education to promote a clear identity of a national society and a regional or professional community.

The introduction of chatbots in messaging systems raises security and privacy concerns, according to article [16]. It emphasizes the possible harm that chatbots bring to users by stealing information from channels without the victim's knowledge. The study provides an approach for automatically assessing security and privacy issues in messaging platform chatbots that involves static and dynamic analysis. The study concentrated on the popular Discord platform and discovered that 55% of chatbots from a top Discord repository sought "administrator" authorization, which could pose a security concern. Furthermore, just 4.35% of chatbots with permissions post a privacy policy. These findings indicate that there are serious security and privacy risks related with the use of chatbots in messaging systems that must be addressed in order to protect users.

The article [17] looked into the drivers of continuation intention in the domain of AI assistants, with a focus on information security. The findings revealed that trust and privacy concerns about one's surroundings are key predictors of intention to continue. This implies that when utilizing AI assistants, users should consider the security of their personal information, and that maintaining trust and privacy are critical considerations for the successful adoption and continuous use of these technologies.

In summary, the topics discussed cover a wide range of chatbots and security issues, such as hostile chatbots, insecure communication channels, authentication and authorisation, data privacy, and social engineering. The research also offered other techniques to dealing with these dangers, such as using end-to-end encryption, including static and dynamic analysis, and putting in place organizational, administrative, and technical controls. The factors of continuation intention in the realm of AI assistants were also investigated, with an emphasis on information security.

Table 2 highlights the research techniques utilized, types of chatbots evaluated, particular security issues addressed, and relevant articles for nine chatbot security studies. The studies studied several sorts of chatbots, such as financial chatbots, instructional chatbots, and messaging platform chatbots, using diverse study methodologies such as case studies, questionnaires, and experimental investigations. Malicious chatbots, unprotected communication channels, authentication and authorization, data privacy, social engineering, and user trust and privacy concerns are among the specific security challenges addressed in the studies. The accompanying articles go into greater detail about the significance of preserving user trust, resolving privacy concerns, and enhancing cyber threat awareness and security. Overall, the table provides a thorough summary of the primary security concerns and weaknesses related with chatbots, as well as the potential solutions.

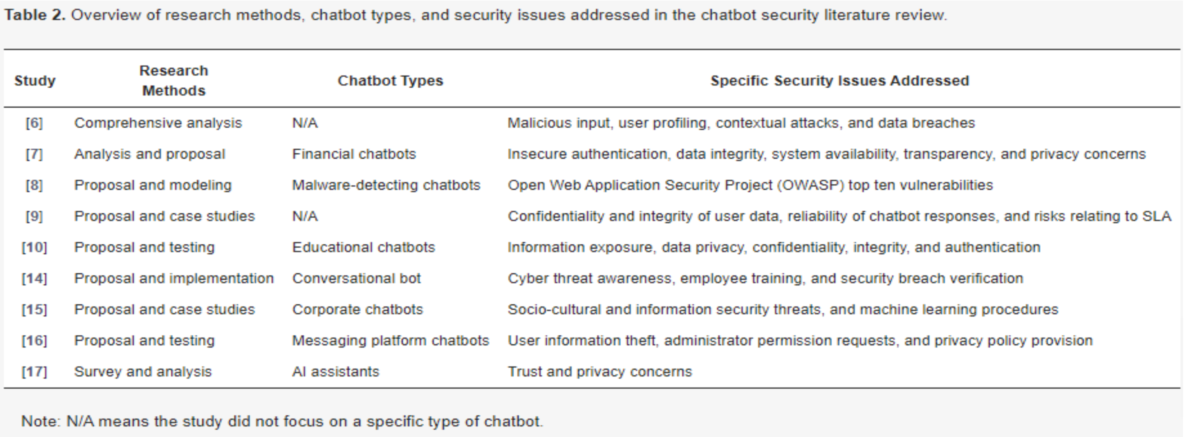


Table 2. Overview of research methods, chatbot types, and security issues addressed in the chatbot security literature review.

Study	Research Methods	Chatbot Types	Specific Security Issues Addressed
[6]	Comprehensive analysis	N/A	Malicious input, user profiling, contextual attacks, and data breaches
[7]	Analysis and proposal	Financial chatbots	Insecure authentication, data integrity, system availability, transparency, and privacy concerns
[8]	Proposal and modeling	Malware-detecting chatbots	Open Web Application Security Project (OWASP) top ten vulnerabilities
[9]	Proposal and case studies	N/A	Confidentiality and integrity of user data, reliability of chatbot responses, and risks relating to SLA
[10]	Proposal and testing	Educational chatbots	Information exposure, data privacy, confidentiality, integrity, and authentication
[14]	Proposal and implementation	Conversational bot	Cyber threat awareness, employee training, and security breach verification
[15]	Proposal and case studies	Corporate chatbots	Socio-cultural and information security threats, and machine learning procedures
[16]	Proposal and testing	Messaging platform chatbots	User information theft, administrator permission requests, and privacy policy provision
[17]	Survey and analysis	AI assistants	Trust and privacy concerns

Note: N/A means the study did not focus on a specific type of chatbot.

4. Taxonomy

Table 3 outlines the prevalent themes and trends observed in the chatbot security literature. The literature emphasizes the significance of recognizing and addressing security threats and vulnerabilities in chatbots. There is also a focus on retaining user confidence and addressing privacy concerns, as well as the need for greater user education and awareness. The significance of safe communication routes and encryption is also emphasized, as is the relevance of AI and machine learning in chatbot security. Standardized security procedures and laws are viewed as vital, yet analyzing and testing chatbot security presents issues. According to the research, there is also a lack of emphasis on insider threats and social engineering attacks.

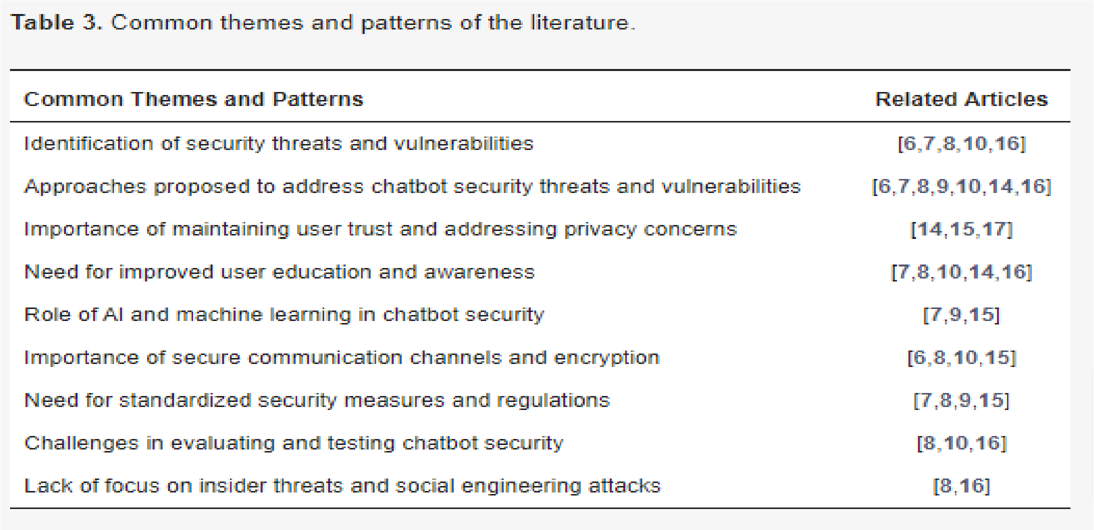
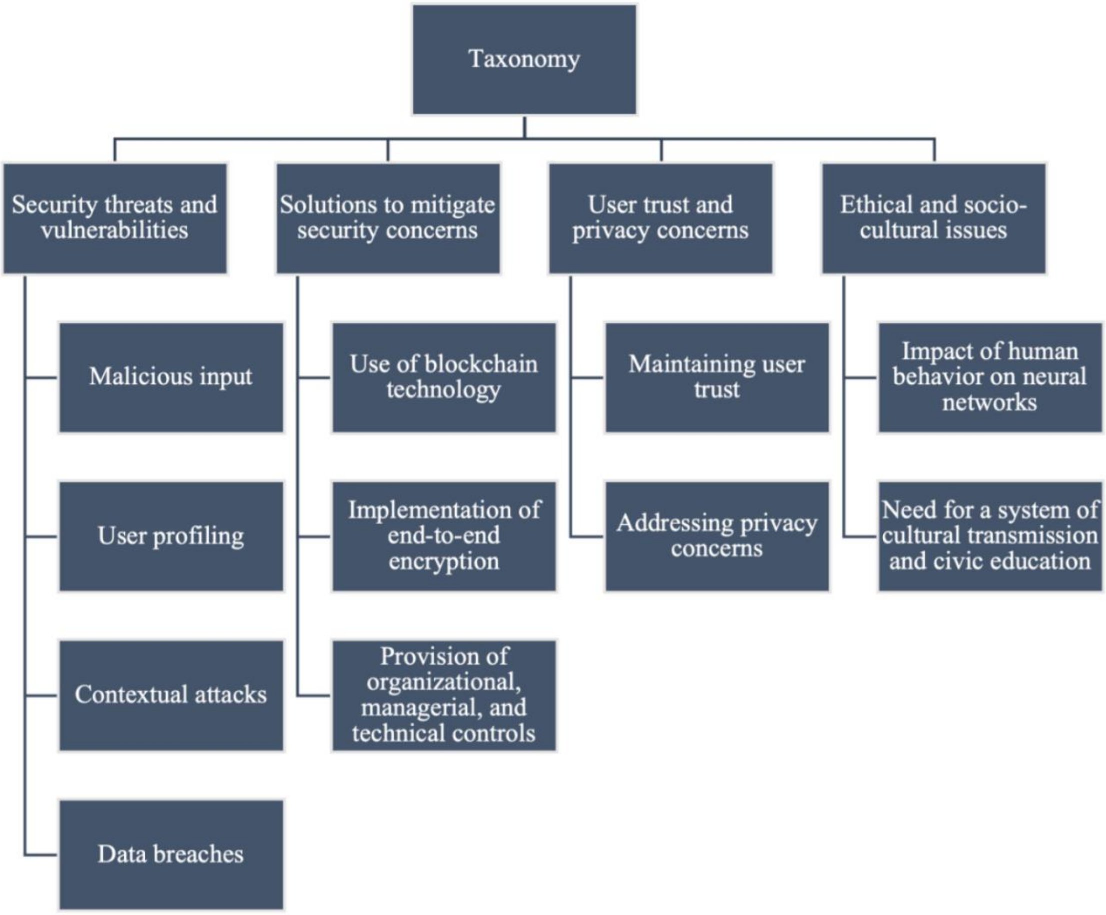


Table 3. Common themes and patterns of the literature.

Common Themes and Patterns	Related Articles
Identification of security threats and vulnerabilities	[6,7,8,10,16]
Approaches proposed to address chatbot security threats and vulnerabilities	[6,7,8,9,10,14,16]
Importance of maintaining user trust and addressing privacy concerns	[14,15,17]
Need for improved user education and awareness	[7,8,10,14,16]
Role of AI and machine learning in chatbot security	[7,9,15]
Importance of secure communication channels and encryption	[6,8,10,15]
Need for standardized security measures and regulations	[7,8,9,15]
Challenges in evaluating and testing chatbot security	[8,10,16]
Lack of focus on insider threats and social engineering attacks	[8,16]

Figure 4 displays the information security taxonomy in chatbots that we created based on the data in Table 3. The taxonomy is designed to provide a framework for understanding the various components of information security in chatbots, as well as to assist identify gaps and areas for additional research. The prevalent themes and patterns observed in the research on chatbot security can be divided into four categories: Themes 1 and 2 address security risks and vulnerabilities, while themes 3 and 4 address user trust and privacy concerns. The first theme addresses four issues: fraudulent input, user profiling, contextual attacks, and data breaches. The usage of blockchain technology, the installation of end-to-end encryption, and the supply of organizational, administrative, and technological controls are all covered in Theme 2. The third theme addresses two issues: retaining user confidence and addressing privacy concerns. The fourth theme addresses two issues: the impact of human conduct on brain networks and the necessity for a cultural transmission and civic education system.


Taxonomy
Security threats and vulnerabilities
Solutions to mitigate security concerns
User trust and privacy concerns
Ethical and socio-cultural issues
Malicious input
User profiling
Contextual attacks
Data breaches
Use of blockchain technology
Implementation of end-to-end encryption
Provision of organizational, managerial, and technical controls
Maintaining user trust
Addressing privacy concerns
Impact of human behavior on neural networks
Need for a system of cultural transmission and civic education

Theme 1: Security threats and vulnerabilities.

The continually developing nature of cyber threats is one of the primary issues in tackling security threats and vulnerabilities in chatbots. New attack vectors and vulnerabilities are constantly being identified, and developers must stay current on the newest security measures to assure their chatbots' continuing protection. The recent increase in phishing attempts utilizing chatbots, for example, emphasizes the significance of installing controls to avoid malicious input and contextual attacks.

Another issue is balancing security with usability. Chatbots must deliver a user-friendly experience while still maintaining user data confidentiality, integrity, and availability. Developers must carefully weigh the trade-offs between security and user experience to strike a balance that fits both objectives. Chatbots in the healthcare industry, for example, must follow tight data privacy and security standards while simultaneously giving fast and accurate medical advice.

Theme 2: Solutions to mitigate security concerns

The lack of defined security procedures for chatbots makes resolving security concerns difficult. Developers must decide which security measures to use, how to use them, and how to assure their effectiveness. End-to-end encryption, for example, is a promising approach for protecting user data, but its implementation may be problematic for small and medium-sized chatbot providers due to financial restrictions.

Another difficulty is the requirement for efficient organizational, managerial, and technical controls. Chatbots are frequently created and deployed by teams with varied levels of information security competence and experience. To prevent security breaches, developers must ensure that all team members receive sufficient information security training and adhere to best practices. Furthermore, developers must ensure that their chatbots are equipped with adequate technical controls, such as access control mechanisms and monitoring tools, to detect and respond to security issues in a timely manner.

Theme 3: User trust and privacy concerns

The lack of transparency in chatbot operations is one of the obstacles in addressing user trust and privacy issues. Users may be hesitant to share sensitive information with chatbots if they are unaware of how their information is being utilized or who has access to it. Chatbot developers must ensure that their chatbots are open about how user data is collected, saved, and used. Chatbots that gather user data for personalized marketing, for example, must provide clear and straightforward information on how the data is used and allow users to opt out if desired.

Another difficulty is addressing cultural and geographical variances in privacy expectations. Expectations and conventions about data privacy and security may differ across cultures and locations. Developers must consider these variations when building and implementing chatbots in order to ensure that they are culturally suitable and respect user privacy.

Theme 4: Ethical and socio-cultural issues

One difficulty in resolving ethical and socio-cultural issues in chatbots is the possibility of unintended repercussions. Chatbots are supposed to interact with humans and learn from such interactions, but their ability to do so can also lead to unintended biases and discrimination. Chatbots trained on biased datasets, for example, may unintentionally propagate prejudices or discriminate against specific groups of users.

Another issue is the necessity for effective cultural transmission and civic education. Chatbots may communicate with users from various cultural and linguistic backgrounds, and developers must guarantee that their chatbots are sensitive to these variations. This may entail including cultural sensitivity training into the development process or collaborating with local organizations to better understand cultural norms and expectations.

Following are a few possible approaches for achieving the security objectives specified in each theme:

Threats and vulnerabilities to security: Developers can conduct regular security audits and vulnerability assessments, implement suitable safety precautions to fix shortcomings, and stay up to date with the latest security threats and vulnerabilities by attending security conferences and training sessions, participating in online forums, and following industry experts on social media to address potential security threats and vulnerabilities in chatbots.

Frequent audits of security and assessments of vulnerability involve analyzing the chatbot system's security measures and discovering any holes or vulnerabilities that cybercriminals might attack. To discover and address potential security concerns, developers can employ a range of tools and techniques, including as a result of penetration testing and vulnerabilities scanning. After holes have been identified, efficient security precautions can be implemented to mitigate them. This could include updating software and firmware to fix known security problems, installing control mechanisms to restrict those with ownership of confidential information, and encryption data in transit and at rest.

It is also critical to stay up to date on the newest security hazards and weaknesses. Security conferences and training sessions are available for developers to learn about emerging threats and best practices for dealing with them. They may also keep up with the newest advances in chatbot security by participating in online forums and following industry experts on social media.

To address security concerns in chatbots, developers can use end-to-end encryption to protect user data from unauthorized access, blockchain technology to improve data security, and organizational, managerial, and technical controls to ensure the confidentiality, integrity, and availability of user data.

End-to-end encryption is a critical security technique that may be applied in chatbots to safeguard user data. This encryption mechanism ensures that the data communicated between the chatbot and the user is safe and cannot be intercepted or viewed by unauthorized third parties. End-to-end encryption protects users' privacy and security while also aiding in the development of confidence in the chatbot system.

End-to-end encryption, for example, may be used by a financial institution's chatbot to safeguard customers' financial information, such as bank account numbers and transaction history. End-to-end encryption encrypts data sent between the user and the chatbot and can only be decrypted by the user or the chatbot, ensuring its confidentiality and security.

The usage of blockchain technology is another security precaution that may be employed. Blockchain is a distributed ledger that may be used to securely store and transfer data. Chatbots may store sensitive data in a decentralized and tamperproof manner using blockchain technology, guaranteeing that it stays safe and cannot be tampered with or altered by unauthorized parties.

A healthcare chatbot, for example, might utilize blockchain to securely store and communicate sensitive patient data such as medical histories and medications. The data may be kept in a decentralized fashion utilizing blockchain technology, making it more resistant to hackers and guaranteeing that patient data remains secure and confidential.

Developers can incorporate organizational, administrative, and technological controls in addition to these technical solutions to secure the security, integrity, and availability of user data. Implementing access restrictions, performing regular security assessments, and offering security awareness training to workers and users are all part of this.

A chatbot utilized by a government organization, for example, may have access restrictions to guarantee that only authorized individuals have access to sensitive data. Regular security audits can also be performed to uncover weaknesses and assure the chatbot system's security. Finally, users can be given security awareness training to assist them understand the importance of data security and how to secure their personal information.

Concerns about user trust and privacy: Developers can address user trust and privacy concerns by maintaining transparency and open communication with users about the data collected and how it is used, obtaining explicit consent from users before collecting or processing their data, and implementing appropriate security measures to protect user data from unauthorized access or disclosure.

Maintaining user trust and managing privacy issues are crucial for chatbot adoption success. Developers can use a variety of ways to do this. For starters, they may guarantee that consumers are adequately informed about the data gathered and how it is utilized. This may be accomplished by making clear and straightforward privacy rules and terms of service available to users. These policies should specify what data is gathered, how it is used, and with whom it is shared.

Second, before collecting or processing data from users, developers can get their express consent. This may be accomplished via a variety of approaches, such as pop-up permission forms, checkboxes, or other interactive mechanisms that explicitly explain why and what data is being gathered. Developers must also guarantee that users have the ability to opt out of data collection or processing if they do not want to disclose their information.

Third, developers can put in place adequate security measures to prevent unauthorized access or exposure of user data. Implementing encryption, access restrictions, and other technical protections to avoid data breaches or leaks is one example. Additionally, developers can make sure that their chatbots abide by pertinent data protection laws and regulations, like the General Data Protection Regulation or the Health Insurance Portability and Accountability Act, and obtain the necessary certifications or third-party audits to prove compliance.

Issues with ethics and culture: To solve these problems, developers should think about how chatbots could affect society and make sure that they are developed and used in an ethical and responsible way. To achieve this, it could be necessary to create a system of civic education and cultural transmission to make sure users are aware of the possible hazards and advantages of chatbots. Additionally, it might be necessary to put in place the proper safety measures to stop chatbots from being used maliciously or unethically.

For instance, if chatbots are not trained to be inclusive and appreciative of diversity, they may reinforce prejudices or preconceptions. Because of the diversity of the consumers they will engage with, developers must make sure that chatbots are created and educated to be culturally aware. Additionally, chatbots may alter the social dynamics of communication, particularly in delicate fields like mental health where they are increasingly employed to help and direct users. Developers must make sure that in these situations, chatbots are enhancing rather than replacing human contact and that consumers have the option to ask for human assistance if necessary.

The potential for harmful usage of chatbots, such as the dissemination of false information or the swaying of public opinion, must also be taken into account by developers. In such circumstances, adequate security measures must be put in place, such as strong authentication systems and restrictions to prevent unauthorized access to chatbots.

Developers should, in general, address chatbot security holistically, taking into account user trust, privacy, and ethical issues in addition to technological security measures. Developers can make sure that their chatbots offer a safe and convenient experience by adhering to best practices and maintaining current with the most recent security threats and vulnerabilities.

5. Discussion

Despite the fact that the examined literature offers a thorough overview of the security risks and vulnerabilities connected to chatbots as well as the methods suggested to solve them, there are still certain places where more study is required to address particular security challenges.

Authentication and authorisation techniques for chatbots are one such topic that need more study. According to article [7], chatbots are vulnerable to unreliable authentication and data integrity problems, which can be used by attackers to obtain unauthorized access to confidential data. Even though certain studies, as article [10], provide methods for safe authentication and encryption procedures, additional study is required to find efficient and scalable solutions that can be applied in many chatbot situations and scenarios.

The identification and prevention of harmful chatbots is another topic that needs more study. There has to be more study on recognizing and detecting rogue chatbots, according to article [3], which explores the potential for malicious chatbots to damage consumers. As mentioned in article [8], which suggests a design and framework for a cyber-aware chatbot service that detects and aids in stopping malware behavior from spreading on the user's machine, more research is required to create effective and efficient methods for monitoring chatbots and spotting suspicious behavior.

Additionally, as stated in article [15], additional study is required to address the security risks and ethical concerns associated with information security in chatbot deployment. More study is specifically required to create frameworks and rules for guaranteeing the privacy, availability, and integrity of user data as well as to address ethical issues around user privacy and data ownership.

Finally, additional study is required to determine how chatbots affect social engineering assaults. According to article [2], chatbots might be used by attackers to carry out social engineering assaults, however further study is required to determine how sensitive chatbots are to these kinds of attacks and to develop effective defenses.

Along with more research, we might also think about some best practices for creating secure chatbots. Best practices for creating secure chatbots include the following:

Conducting thorough security evaluations throughout the whole development process for chatbots is one of the best practices for creating safe chatbots. Before deploying the chatbot, developers should carry out security testing at various stages to find and fix flaws. These evaluations must to encompass every component of the chatbot architecture, including the client, communication, response generating, and database modules. A developer may, for instance, utilize static analysis tools to find code-level vulnerabilities like buffer overflows or SQL injections and dynamic analysis tools to test for vulnerabilities while the program is running.

Implementing user authentication and authorization is another recommended practice for creating safe chatbots. The chatbot should require users to verify themselves before accessing sensitive data or services, according to the developers. Traditional username and password combinations, two-factor authentication, biometric authentication, or autonomous inquiry-based authentication are all possible methods of authentication [32]. In addition, authorisation should be used to guarantee that users can only access the information or services to which they have been granted access. For instance, customers should only be able to access their own accounts using a chatbot used for banking.

Developers should use encryption to safeguard sensitive data in chatbots. Data transmission may be made more secure via encryption by preventing unauthorized parties from intercepting and viewing it. Developers can employ authenticated key agreement systems [33,34,35] in various contexts to secure data in transit and transport layer security or symmetric or asymmetric encryption to secure data saved on the chatbot's database.

Chatbots should be frequently updated and patched to address issues and improve privacy [36]. Developers should keep an eye out for notifications and advisories, and keep their chatbots up to current with the most recent updates for safety. Chatbots should also be monitored for unusual actions, and logs should be reviewed on a regular basis to discover any security breaches.

6. Conclusions

Chatbot information security is a critical aspect of any chatbot implementation. As chatbots interact with users, they handle sensitive and personal information, making them potential targets for security breaches. Therefore, a systematic analysis of chatbot information security is essential to identify potential vulnerabilities and establish robust security measures.

Chatbots should implement secure authentication mechanisms to verify the identity of users before granting access to sensitive information. This may involve techniques like password-based authentication, multi-factor authentication, or integration with existing authentication systems.

To protect the confidentiality of user data, chatbots should employ encryption techniques when storing or transmitting sensitive information. Encryption ensures that even if data is intercepted, it remains unreadable without the decryption key.

Chatbots should utilize secure communication protocols, such as HTTPS (HTTP over SSL/TLS), to encrypt the data exchanged between the chatbot and the user. This helps prevent eavesdropping and data tampering during transmission.

Chatbots should validate and sanitize user input to prevent common security vulnerabilities like SQL injection, cross-site scripting (XSS), and command injection. Implementing strict input validation routines can help mitigate these risks. Chatbots should adhere to privacy regulations and protect user privacy. They should clearly communicate their data collection and usage policies and obtain user consent where necessary. Additionally, anonymization techniques can be employed to minimize the risk of exposing personally identifiable information. Proper error handling and logging mechanisms should be implemented to capture and monitor potential security incidents or abnormal activities. Logs should be securely stored and regularly reviewed to identify any suspicious patterns or unauthorized access attempts. Implementing access controls helps ensure that only authorized individuals or systems can interact with and manage the chatbot. This includes controlling administrative access, restricting permissions, and using role-based access control (RBAC) mechanisms. Chatbot systems should be regularly updated with security patches and fixes to address any identified vulnerabilities. Regular vulnerability assessments and penetration testing can help identify weaknesses and ensure appropriate remediation.

References

1. Dhinagaran, D.A.; Martinengo, L.; Ho, M.-H.R.; Joty, S.; Kowatsch, T.; Atun, R.; Car, L.T. Designing, Developing, Evaluating, and Implementing a Smartphone-Delivered, Rule-Based Conversational Agent (DISCOVER): Development of a Conceptual Framework. *JMIR Mhealth Uhealth* **2022**, *10*, e38740. [**Google Scholar**] [**CrossRef**] [**PubMed**]
2. Adamopoulou, E.; Moussiades, L. An Overview of Chatbot Technology. In Proceedings of the Artificial Intelligence Applications and Innovations 2020, Neos Marmaras, Greece, 5–7 June 2020. [**Google Scholar**]
3. Adamopoulou, E.; Moussiades, L. Chatbots: History, technology, and applications. *Mach. Learn. Appl.* **2020**, *2*, 100006. [**Google Scholar**] [**CrossRef**]
4. Chen, C.-M.; Liu, S.; Li, X.; Islam, S.H.; Das, A.K. A provably-secure authenticated key agreement protocol for remote patient monitoring IoMT. *J. Syst. Arch.* **2023**, *136*, 102831. [**Google Scholar**] [**CrossRef**]
5. Chen, C.-M.; Li, Z.; Kumari, S.; Srivastava, G.; Lakshmanna, K.; Gadekallu, T.R. A provably secure key transfer protocol for the fog-enabled Social Internet of Vehicles based on a confidential computing environment. *Veh. Commun.* **2023**, *39*, 100567. [**Google Scholar**] [**CrossRef**]
6. Ye, W.; Li, Q. Chatbot Security and Privacy in the Age of Personal Assistants. In Proceedings of the 2020 IEEE/ACM Symposium on Edge Computing, San Jose, CA, USA, 12–14 November 2020. [**Google Scholar**]
7. Bhuiyan, M.S.I.; Razzak, A.; Ferdous, M.S.; Chowdhury, M.J.M.; Hoque, M.A.; Tarkoma, S. BONIK: A Blockchain Empowered Chatbot for Financial Transactions. In Proceedings of the 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications, Guangzhou, China, 29 December 2020–1 January 2021. [**Google Scholar**]
8. Thorpe, S.; Scarlett, H. Towards a Cyber Aware Chatbot Service. In Proceedings of the 2021 IEEE International Conference on Big Data, Orlando, FL, USA, 15–18 December 2021. [**Google Scholar**]
9. Gondaliya, K.; Butakov, S.; Zavarsky, P. SLA as a mechanism to manage risks related to chatbot services. In Proceedings of the 2020 IEEE 6th Intl Conference on Big Data Security on Cloud, IEEE Intl Conference on High Performance and Smart Computing, and IEEE Intl Conference on Intelligent Data and Security, Baltimore, MD, USA, 25–27 May 2020. [**Google Scholar**]
10. Shah, M.; Panchal, M. Privacy Protected Modified Double Ratchet Algorithm for Secure Chatbot Application. In Proceedings of the 2022 3rd International Conference on Smart Electronics and Communication, Trichy, India, 20–22 October 2022. [**Google Scholar**]
11. Belen-Saglam, R.; Nurse, J.R.C.; Hodges, D. An Investigation Into the Sensitivity of Personal Information and Implications for Disclosure: A UK Perspective. *Front. Comput. Sci.* **2022**, *4*, 1–22. [**Google Scholar**] [**CrossRef**]
12. Patil, K.; Kulkarni, M.S. Artificial intelligence in financial services: Customer chatbot advisor adoption. *Int. J. Innov. Technol. Explor. Eng.* **2019**, *9*, 4296–4303. [**Google Scholar**] [**CrossRef**]
13. Ali, H.; Aysan, A.F. What will ChatGPT Revolutionize in Financial Industry? *Soc. Sci. Res. Netw.* **2023**, 4403372. [**Google Scholar**] [**CrossRef**]
14. El Hajal, G.; Daou, R.A.Z.; Ducq, Y. Human Firewall: Cyber Awareness using WhatApp AI Chatbot. In Proceedings of the 2021 IEEE 3rd International Multidisciplinary Conference on Engineering Technology, Beirut, Lebanon, 8–10 December 2021. [**Google Scholar**]

15. Pokrovskaia, N.N. Sociocultural and Information Security Issues in the Implementation of Neural Network Technologies in Chat-bots Design. In Proceedings of the 2022 XXV International Conference on Soft Computing and Measurements, Saint Petersburg, Russia, 25–27 May 2022. [**Google Scholar**]
16. Edu, J.; Mulligan, C.; Pierazzi, F.; Polakis, J.; Suarez-Tangil, G.; Such, J. Exploring the security and privacy risks of chatbots in messaging services. In Proceedings of the 22nd ACM Internet Measurement Conference, Nice, France, 25–27 October 2022. [**Google Scholar**]
17. Jo, H. Impact of Information Security on Continuance Intention of Artificial Intelligence Assistant. *Procedia Comput. Sci.* **2022**, *204*, 768–774. [**Google Scholar**] [**CrossRef**]
18. Nadarzynski, T.; Miles, O.; Cowie, A.; Ridge, D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *Digit. Health* **2019**, *5*, 1–12. [**Google Scholar**] [**CrossRef**]
19. Waheed, N.; Ikram, M.; Hashmi, S.S.; He, X.; Nanda, P. An Empirical Assessment of Security and Privacy Risks of Web-Based Chatbots. In Proceedings of the International Conference on Web Information Systems Engineering, Biarritz, France, 1–3 November 2022. [**Google Scholar**]
20. Hasal, M.; Nowaková, J.; Saghair, K.A.; Abdulla, H.; Snášel, V.; Ogiela, L. Chatbots: Security, privacy, data protection, and social aspects. *Concurr. Comput. Pract. Exp.* **2021**, *33*, 1–13. [**Google Scholar**] [**CrossRef**]
21. Følstad, A.; Nordheim, C.B.; Bjørkli, C.A. What makes users trust a chatbot for customer service? An exploratory interview study. In Proceedings of the International Conference on Internet Science, St. Petersburg, Russia, 24–26 October 2018. [**Google Scholar**]
22. van der Goot, M.J.; Pilgrim, T. Exploring Age Differences in Motivations for and Acceptance of Chatbot Communication in a Customer Service Context. In Proceedings of the International Workshop on Chatbot Research and Design, Amsterdam, The Netherlands, 19–20 November 2019. [**Google Scholar**]
23. United Nations, Department of Economic and Social Affairs, Population Division. Available online: **http://esa.un.org/wpp/** (accessed on 12 May 2023).
24. GPT-4 Is OpenAI’s Most Advanced System, Producing Safer and More Useful Responses. Available online: **https://openai.com/product/gpt-4** (accessed on 12 May 2023).
25. Corsello, A.; Santangelo, A. May Artificial Intelligence Influence Future Pediatric Research?—The Case of ChatGPT. *Children* **2023**, *10*, 757. [**Google Scholar**] [**CrossRef**] [**PubMed**]
26. Kooli, C. Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability* **2023**, *15*, 5614. [**Google Scholar**] [**CrossRef**]
27. Giansanti, D. The Chatbots Are Invading Us: A Map Point on the Evolution, Applications, Opportunities, and Emerging Problems in the Health Domain. *Life* **2023**, *13*, 1130. [**Google Scholar**] [**CrossRef**]
28. Aggarwal, A.; Tam, C.C.; Wu, D.; Li, X.; Qiao, S. Artificial Intelligence–Based Chatbots for Promoting Health Behavioral Changes: Systematic Review. *J. Med. Internet Res.* **2023**, *25*, e40789. [**Google Scholar**] [**CrossRef**]
29. Sallam, M. ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare* **2023**, *11*, 887. [**Google Scholar**] [**CrossRef**]
30. 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.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Syst. Rev.* **2021**, *10*, 89. [**Google Scholar**] [**CrossRef**]
31. Tarsilla, M. Cochrane Handbook for Systematic Reviews of Interventions. *J. Multidiscip. Eval.* **2010**, *6*, 143–148. [**Google Scholar**] [**CrossRef**]
32. Voege, P.; Abu Sulayman, I.I.M.; Ouda, A. Smart Chatbot for User Authentication. *Electronics* **2022**, *11*, 4016. [**Google Scholar**] [**CrossRef**]
33. Wu, T.-Y.; Meng, Q.; Chen, Y.-C.; Kumari, S.; Chen, C.-M. Toward a secure smart-home IoT access control scheme based on home registration approach. *Mathematics* **2023**, *119*, 2123. [**Google Scholar**] [**CrossRef**]
34. Wu, T.-Y.; Kong, F.; Meng, Q.; Kumari, S.; Chen, C.-M. Rotating Behind Security: An enhanced authentication protocol for IoT-enabled devices in distributed cloud computing architecture. *EURASIP J. Wirel. Commun. Netw.* **2023**, *2023*, 36. [**Google Scholar**] [**CrossRef**]
35. Wu, T.-Y.; Meng, Q.; Yang, L.; Kumari, S.; Pirouz, M. Amassing the Security: An Enhanced Authentication and Key Agreement Protocol for Remote Surgery in Healthcare Environment. *Comput. Model. Eng. Sci.* **2023**, *134*, 317–341. [**Google Scholar**] [**CrossRef**]
36. Chow, J.C.; Sanders, L.; Li, K. Design of an educational chatbot using artificial intelligence in radiotherapy. *AI* **2023**, *4*, 319–332. [**Google Scholar**] [**CrossRef**]

37. Galhotra, B., & Lowe, D. (2023). Utilizing digital technology for chatbot innovation in e-business. *International Journal of Innovative Research in Engineering & Management, 10*(5), 1-5. https://doi.org/10.55524/ijirem.2023.10.5.1

38. Sharma, D., Saxena, A. B., & Aggarwal, D. (2024). Harnessing AI chatbots for personalized learning: Implications and strategies for education. Journal of Informatics Education and Research, 4(2), 1-10. http://jier.org