

An investigation on strategies for optimizing consumer trust in chatbots

Xi Ning Luo

Faculty of Arts and Science, University of Toronto, Toronto, M5R0A3, Canada

cynth.luo@mail.utoronto.ca

Abstract. The advancement of artificial intelligence (AI) gave rise to chatbots, which is a type of AI-powered software that communicates via natural language. Chatbots have been used in diverse contexts, delivering significant convenience to the consumers. Nonetheless, this technology encounters ambivalent attitudes from consumers. Some aspects of the chatbot technology are evoking distrustful attitudes among consumers, while the others are cultivating a sense of trust. Thus, the objective of the current paper is to outline and analyze key factors that affect consumer trust and elucidate strategies that firms can adopt to optimize trust. According to recent studies, consumer distrust primarily stems from algorithmic bias, privacy and security concerns, and the lack of algorithmic transparency; on the other hand, consumer trust is formed due to anthropomorphic attributes of chatbots, particularly warmth and competence. To reduce consumer distrust, companies are advised to first identify and minimize existing real risks in their products, then deliver transparency to the public to establish a trustworthy image. To increase trust, companies are suggested to improve upon the anthropomorphic attributes of chatbots. Contributions and limitations of the paper are also discussed to highlight areas that require further investigation in the field of chatbots as well as AI in general.

Keywords: Artificial Intelligence, Chatbot, Trust, Algorithmic Transparency, Anthropomorphism.

1. Introduction

The advent of the age of Big Data has propelled the advancement of artificial intelligence (AI), a technology that heavily involves data manipulation and analysis. AI is defined as computational agents constructed using algorithmic models, with the aim to imitate human capabilities while exceeding humans in accuracy [1]. Currently, AI has a variety of implementations across society, with interactive products and services standing out as they exert the most direct influence on the general public. One such type of product is chatbot.

The chatbot is an AI-powered software that is able to maintain conversations through natural languages. To produce ideal outcomes, the AI within the chatbot needs to have access to vast amounts of data, which are essentially collected from the consumers [2]. However, along with the collection of consumers' personal information emerged serious problems. Public awareness regarding transparency, privacy, and security issues is increasing, resulting in growing distrust towards chatbots; at the same time, companies are facing dilemmas in meeting the public's demand for algorithmic transparency due to concerns about sensitive information in the data collected [3].

The chatbot is a technology with tremendous potential. With the advancement of AI, the public's reliance on chatbots is continuously and inevitably increasing. Hence, this paper aims to investigate key factors that contribute to both consumer trust and distrust in chatbots, along with the underlying challenges in eliminating distrust. Moreover, it will provide insights to assist relevant firms in optimizing consumer trust in chatbots.

2. Chatbot

2.1. Overview of Chatbot Usage

AI conversational agents, commonly referred to as chatbots, are software applications that are capable of engaging in quick and direct human-computer interactions through natural language communication [2]. The AI technology that chatbots are constructed based on enables them to mimic human behavior during conversations. Currently, chatbots are implemented in a variety of contexts, including entertainment, marketing, education, health care, support systems, and cultural diffusion, providing assistance to the public by answering questions that are difficult to find and saving time [4, 5].

2.2. Inner Workings of a Chatbot

A chatbot is implemented using sets of algorithms, specifically employing a technique of AI known as “pattern matching”, which requires a database from which the chatbot can select the most matching responses for any user input [5]. The production of each response for a given input follows a specific procedure. First, a set algorithm examines the input message and analyzes huge amounts of relevant data from the database to detect useful patterns. Then, another set of algorithms generates potential responses based on the database and evaluates the relevance of each response. Finally, a response with the highest perceived relevance is selected and output to the user.

3. Consumer Distrust

3.1. Algorithmic Bias

There is a general consensus that machines are always logical and objective, since they process information purely through algorithms. However, it is crucial to recognize that both the information provided to the machines and the algorithms that machines rely upon originate from human sources, meaning that any existing biases in human society will eventually be reflected in the outcomes produced by machines, and chatbots are no exception.

Algorithmic bias describes the repetition and reinforcement of existing biases in the human world by machines. This phenomenon is primarily a consequence of data bias and method bias. Data biases arise when the training dataset for the algorithm is inadequate or not representative of an arbitrary sample from the target population. As a result, the algorithm is unable to make appropriate decisions about groups that are underrepresented in the dataset, ultimately resulting in bias. Method biases occur when the methods used for training are embedded with biases from the developers, who may favor information that aligns with their own beliefs. Consequently, this leads to correlation fallacy or overgeneralization of findings that are only applicable under specific circumstances [6].

Algorithmic bias can reinforce long-standing prejudices and inequalities in our society, exacerbating the existing challenges faced by already disadvantaged groups, such as discriminatory pricing and restricted access to resources [6]. Given these harmful impacts, algorithmic bias in chatbots can lead to questioning, fear, and dissatisfaction among the general public, in particular, marginalized groups; this adversely affects the overall perceived quality of information provided by chatbots and the enjoyment from using chatbots. Alagarsamy and Mehrolia have aimed to investigate the effect of various factors on consumer trust in chatbots through a study conducted on users of banking chatbot services. The data was collected from an online questionnaire and respondents were chosen through social media platforms. The study results demonstrate that there is a positive relationship between consumer trust and perceived

quality of information provided by chatbots, as well as between consumer trust and the level of enjoyment during chatbot usage [7].

Based on these findings, it is evident that algorithmic bias is a key factor in causing consumer distrust in chatbots. Furthermore, given that consumers do not always have full access to AI algorithms, the potential presence of algorithmic bias in AI products increases consumers' perceived risk of algorithmic bias in chatbots, which is also detrimental to consumer trust.

3.2. Privacy and Security Concerns

In order to optimize the appropriateness of the produced responses, a chatbot needs to have access to databases that are adequate and up-to-date. This essentially requires companies to constantly collect contemporary data from the external world, including personal and demographic information of the general public.

However, the collection of data raises problems related to consumers' privacy. First, there may exist flaws in the computer system used for data storage, which can be exploited by hackers to steal consumer data for unethical uses [7]. Second, since it is difficult to trace harms inflicted on consumers back to their origin of data misuse, the service providers do not need to, and therefore, might not fully internalize potential pitfalls of the algorithms. Third, as it is hard to detect harm inflicted on consumers due to data misuse, the service providers might go back on their consumer-friendly data policy. Finally, the service providers might have more information about future use of consumers' personal data than the consumers themselves; consequently, some consumers hesitate to give away their personal information and must trade-off between personal security and potential loss from beneficial use of their personal data in the future [8].

These issues suggest that there is an unfair bargain of data use between chatbot service providers during personal data collection. Such unfairness increases the consumers' tendency to believe that their potential loss outweighs their potential gain from donating personal information. As a result, the perceived risk—which is specifically defined as a person's subjective evaluation of incurring losses from using a product [7]—of chatbots increases among consumers. Alagarsamy and Mehroliya have identified that there is a negative relationship between consumer trust and perceived risk of chatbots. This outcome can be explained by the fact that most consumers do not know exactly how their personal information is handled by chatbot service providers [7]. Thus, it is evident that any perceived privacy or security concerns about chatbots negatively impact consumer trust.

3.3. Dilemmas in Delivering Algorithmic Transparency

Algorithmic transparency means high accessibility and explainability of data and algorithms utilization as well as operation of the systems. Having algorithmic transparency in an AI system means that any user should be able to understand the functioning of the systems and the rationale behind their utilization of data and algorithms [9]. Grimmelikhuijsen has demonstrated that both the accessibility and explainability of algorithms exert a positive influence on consumers' perceived trustworthiness of the algorithms through survey experiments [10]. As major factors of consumer distrust toward AI systems (such as perceived risk of algorithmic bias and perceived privacy risks) primarily stem from information asymmetries between consumers and service providers, this finding suggests that algorithmic transparency is an effective way of overcoming these asymmetries, ultimately fostering consumer trust in AI products such as chatbots.

To ameliorate consumer trust in chatbots, it is necessary for companies to deliver algorithmic transparency. Nevertheless, some issues are hindering companies during the process of fulfilling algorithmic transparency. Firstly, different levels of transparency may be required depending on whom the AI system is intended to be transparent for. In order to optimize consumer satisfaction, companies should grant individuals the most helpful and relevant information, rather than overwhelming them with exhaustive details about every single aspect. Secondly, the databases that chatbots rely upon very likely contain sensitive and private data about consumers. Delivering transparency may expose this sensitive information, particularly during the algorithms' training phase, thus resulting in undermining of privacy

rights. Thirdly, being able to describe the inner workings of an AI system may not be the equivalent of being able to comprehend and control it. It must be recognized that AI systems are becoming increasingly complex. In some cases, the instructions sent to the system are unsupervised by the developers and therefore can be unintelligible by humans [3].

These issues present significant challenges for chatbot companies in meeting the consumers' demand for algorithmic transparency. Hence, many companies on the market are yet to have the appropriate degree of transparency for their intended consumers. As a result, this deficiency in algorithmic transparency leads to further consumer distrust of many chatbot products and services.

4. Consumer Trust From the Perspective of Anthropomorphism

Despite being an AI system by nature, chatbots interact with users through natural languages, which are products of human society. Chatbots' proficient use of natural languages gives human users a sense of familiarity, reminds them of human contact and friendliness, and stimulates their social presence. This phenomenon of chatbots resembling humans to some extent is referred to as anthropomorphism [11].

Currently, the main anthropomorphic traits of chatbots are warmth and competence. Warmth comprises qualities associated with universal benevolence in human society, such as friendliness, kindness, sincerity, and compassion. On the other hand, competence encompasses qualities that are conventionally regarded as pragmatic and utilitarian, such as skillfulness, knowledgeableness, and efficiency [12]. Past studies have shown that these anthropomorphic characteristics in chatbots exert a positive influence on consumer trust. Cheng and their colleagues investigated the influence of anthropomorphic attributes on consumers' trust in chatbots by employing interviews with consumers who had experience interacting with text-based chatbots. The study results reveal that perceived warmth and perceived competence both contribute to higher trust in chatbots [12]. Hsiao and Chen conducted a survey study with the same objective and identified that chatbots' problem-solving skills and ability to evoke positive emotions enhance their anthropomorphism, thereby increasing consumer trust [11].

These two aspects of anthropomorphism in chatbots serve as fundamental factors contributing to the establishment of trust among consumers. They achieve this by forming a sense of social proximity between consumers and machines, although through different mechanisms.

4.1. Influence of Anthropomorphic Warmth

Experiencing warmth during social interaction gives people a sense of social proximity and evokes positive emotions within them, thus increasing their perceived value and trustworthiness of the other party. In the context of interacting with a chatbot, when the user receives warmth during the interaction, they experience positive feelings such as excitement, happiness, and satisfaction. These potential positive emotional responses can be collectively summarized as the perceived enjoyment. Positive emotional experience from using chatbots reminds consumers of human benevolence. Similar to the emotional dynamics observed during social interactions among humans, this causes consumers to regard chatbots as more valuable and trustworthy. As a result, consumers place a higher level of trust in chatbots [7].

4.2. Influence of Anthropomorphic Competence

When another human being is able to solve simple or complex tasks easily, people will acknowledge the capability of that particular individual and perceive them as trustworthy. In the context of chatbot interaction, a chatbot successfully helps the user when its response aligns with the user's intended purpose for using it. Major factors that contribute to the success rate of helping the user include information quality and service quality. Information quality measures the relevance, completeness, understandability, and accuracy of the response provided by the chatbot. Service quality measures the responsiveness, reliability, and personalization capability of the chatbot. In addition, studies suggest that information quality affects service quality. When a chatbot is deemed as having excellent information quality or service quality, it achieves a high success rate in helping the consumers, reminding them of the competence they experience during interactions with other humans. Similar to the dynamics that

occur when humans help one another, this will make the consumers regard the chatbot as more trustworthy [7].

5. Discussion

The objective of the current paper is to investigate key factors that contribute to both consumer trust and distrust in chatbots and provide insights to assist relevant firms in optimizing consumer trust in chatbots. In order to increase consumer trust in chatbots, firms should not only amplify the impact of existing factors that foster trust, but also mitigate the influence of factors that undermine trust.

In summary, consumers view chatbots as untrustworthy primarily due to potential risks of AI systems. These potential risks can be dissected into two aspects, namely actual risk and perceived risk. Actual risks are tangible, concrete hazards of chatbots that have the potential to inflict real harm upon consumers; these include algorithmic bias, as well as the companies' unreliability and dishonesty in handling consumer data. Conversely, perceived risks are suspicions regarding potential hazards arising from the reporting of actual risks; these include perceived algorithmic bias and perceived privacy risks. Since perceived risks are formed as a result of the existence of actual risks, companies should place a priority on eliminating actual risks.

The first major actual risk is algorithmic bias. This type of risk can be categorized into data bias and method bias, and each of these categories requires different mitigation strategies. The sources of data bias manifest during data collection and data preparation. To address these sources, it is suggested for companies to thoroughly document the origin and utilization of all datasets. These documents enable managers to thoroughly assess all potential risks throughout the creation and maintenance of a dataset, thus effectively minimizing the occurrence of data bias. On the other hand, method bias originates from methodological issues that stem from biases exhibited by the developers. To resolve these issues, it is suggested for companies to create model cards, which are documents that contain key information about algorithmic models. Model cards enable engagement within the developers as well as other key stakeholders, allowing them to more effectively identify and eliminate methodological biases [6].

The second major actual risk is chatbot companies' unreliability and dishonesty in handling consumer data. To address this risk, companies should first maximize the security of their database systems, as this measure can directly reduce the likelihood of data misuse from a third party. Aspects that companies are suggested to consider to bolster cybersecurity include revising insecure coding, upgrading out-of-date hardware drivers, and selecting stronger firewalls [7]. Next, companies should minimize data misuse from themselves. As discussed previously, companies should preserve detailed documents for every stage of data handling. These documents can assist managers in detecting data misuse within the company, reducing the difficulty of tracing data misuse to its origin. This subsequently prevents deviations from consumer-friendly data policy and increases the company's overall integrity.

To address perceived risks, chatbot companies need effective delivery of algorithmic transparency, as it informs the consumers that their chatbots are safe and secure. To achieve this, companies should first determine their core consumer group and its primary demands, then identify information that are relevant to these demands, and lastly select the appropriate method for disclosing the information. It is important to ensure that the information publicized is concise, as overwhelming the consumers with irrelevant information may backfire by decreasing their satisfaction [3]. Moreover, companies should also prevent AI over appreciation, which is the phenomenon that consumers regard AI systems as so superior that they uncritically rely on it. AI over appreciation is harmful, as it inhibits the public from recognizing potential flaws and errors of AI systems [13]. Thus, when disclosing key information, companies should demonstrate both the advantages and disadvantages of their products and maintain a neutral attitude. Showcasing the advantages of a company's product ameliorate the trust of its consumers, while acknowledging the disadvantages prevents AI over appreciation.

On the other hand, companies should also amplify the effect of anthropomorphic attributes of chatbots to boost consumer trust. First, companies can make the responses produced by chatbots more human-like in general, such as employing emoticons or stickers [11]. Second, companies are advised to identify the type of relationship norm that their consumers demand the most and adjust the chatbots

accordingly. Depending on the consumers' most prominent desire, they will place more weight on the chatbot's anthropomorphic warmth or anthropomorphic competence [12]. If the consumers wish to establish an exchange relationship with chatbots, they will value the chatbot's competence more than its warmth; conversely, if the consumers are seeking an emotionally comfortable relationship, they will emphasize more on its warmth. If a chatbot achieves superior performance in its most demanded aspect, it subsequently gains more trust from consumers.

6. Conclusion

The chatbot is a currently advancing technology with the ability to provide unprecedented convenience to the public. With the surge of AI, people's reliance on chatbots will inevitably and continuously increase. Therefore, it is crucial to establish consumer trust in chatbots, not only for providing better user experience and public opinion, but also for facilitating firms in optimizing the future potential of this technology.

The paper offers several theoretical strategies that companies can consider and adopt to improve consumer trust, mitigate consumer distrust, and minimize the harm inflicted upon consumers. Firstly, the paper outlines approach that companies can take to minimize existing risks in their chatbot product. Secondly, although algorithmic transparency has been proven to be crucial in gaining consumer trust, the paper highlights the pitfalls of algorithmic transparency if not delivered correctly; this underscores the importance for companies to carefully examine their information to be delivered as well as method of delivery. Thirdly, the paper identifies that the anthropomorphic attributes within chatbots are key to gaining consumer trust and provides general insights on enhancing these attributes.

Furthermore, this paper contributes to the future research of chatbot as well as AI. Firstly, through summarizing and analyzing the influential risk factors that reinforce consumer distrust toward chatbots, the paper is able to reason that these risks arise from the AI system that chatbots rely upon and categorize them into actual risk and perceived risk. Secondly, the paper offers a glimpse on AI over appreciation as a result of incorrect ways of delivering algorithmic transparency. Both of these perspectives introduce new avenues for future research into consumer perceptions, attitudes, and behaviors, extending beyond chatbots to encompass other AI-powered products.

Nevertheless, the paper still has limitations. The proposals in this paper are entirely theoretical, since they are not substantiated by any studies. Thus, several directions are suggested for future studies in relevant fields. First, researchers could investigate the relationship between consumers' perceived risk and the actual level of risk associated with AI products under various circumstances. Second, researchers could examine the difference between various forms of information disclosure on consumer perception of AI products.

References

- [1] N. Ameen et al. "Customer experiences in the age of artificial intelligence". *Computers in Human Behavior*, vol. 114, Jan. 2021.
- [2] R. Benabdelouahed & C. Dakouan. "The Use of Artificial Intelligence in Social Media: Opportunities and Perspectives". *Expert Journal of Marketing*, vol. 8, pp. 82-87, 2020.
- [3] LUISS. "Algorithmic Transparency Between Legal and Technical Issues". Available: http://tesi.luiss.it/30511/1/230231_DI%20TORO_GIOVANNA.pdf, 2021. [Accessed: Sep. 13, 2023].
- [4] L. Nicolescu & M. T. Tudorache. "Human-Computer Interaction in Customer Service: The Experience with AI Chatbots—A Systematic Literature Review". *Electronics*, vol. 11, pp. 1579, May 2022.
- [5] M. Dahiya. "A Tool of Conversation: Chatbot". *International Journal of Computer Sciences and Engineering*, vol. 5, pp. 158-161, May 2017.
- [6] S. Akter et al. "Algorithmic bias in data-driven innovation in the age of AI". *International Journal of Information Management*, vol. 60, Oct. 2021.

- [7] S. Alagarsamy & S. Mehroliia. “Exploring chatbot trust: Antecedents and behavioural outcomes”. *Heliyon*, vol. 9, May 2023.
- [8] L. Abrardi, C. Cambini, & L. Rondi. “Artificial intelligence, firms and consumer behavior: A survey”. *Journal of Economic Surveys*, vol. 36, pp. 969-991, Sep. 2022.
- [9] J. Bang et al. “Ethical Chatbot Design for Reducing Negative Effects of Biased Data and Unethical Conversations”. 2021 International Conference on Platform Technology and Service (PlatCon), pp. 1-5, 2021.
- [10] S. Grimmelikhuijsen. “Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-Making”. *Public Administration Review*, vol. 83, pp. 241-262, Feb. 2022.
- [11] K. Hsiao & C. Chen. “What drives continuance intention to use a food-ordering chatbot? An examination of trust and satisfaction”. *Library Hi Tech*, vol. 40, pp. 929-946, 2022.
- [12] X. Cheng et al. “Human vs. AI: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms”. *Information Processing & Management*, vol. 59, May 2022.
- [13] J. Zerilli et al. “How transparency modulates trust in artificial intelligence”. *Patterns*, vol. 3, pp. 1-10, Apr. 2022.