

Trust in the system and human autonomy in customer service chatbots

Ana Gervazoni¹, Manuela Quaresma¹

anagervazoni@tecgraf.puc-rio.br, mquaresma@puc-rio.br

¹Pontifícia Universidade Católica do Rio de Janeiro, Brazil

Abstract

When Artificial Intelligence systems are not explained clearly to users, it can negatively affect their interactions and compromise their perceptions of a brand. When designing and developing conversational agents that deal with the client, it is crucial to consider that they are a service and follow human-centered Artificial Intelligence (HCAI) approaches. This study discusses two HCAI frameworks, relate them to trust in the system and human autonomy and define how these guidelines could be met in customer service chatbot. A survey was conducted to determine if users' views about their interactions with chatbots aligned with the recommended guidelines and how this affected their senses mentioned above. The analysis of the responses indicates that those human-centered Artificial Intelligence approaches still need to be prioritized or even met in customer service chatbot development. Users have reported unpleasant experiences with such services, leading to a decrease in their trust and autonomy.

Keywords: artificial intelligence, service design, chatbot, customer relationship

Introduction

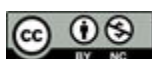
The evolution and diffusion of artificial intelligence (AI) techniques give rise to a new operating model for companies. The automation of internal processes, whether strategic, tactical, or operational, enables greater efficiency at different moments in the production chain.

The relationship between consumers and companies, for a long time carried out by real people answering phones, sending emails or replying messages, has changed drastically over the last few years. Through various communication channels, the customer now gets in touch with robots that mimic human agents, and chatbots are among the most popular of these.

Chatbots are interactive applications capable of communicating with users using Natural Language Processing (NLP) and, starting in 2016, advances in the field of AI allowed companies to start developing their own conversational agents. (Nicolescu and Tudorache, 2022). According to Adamopoulou and Moussiades (2020), “chatbots have become so common because they reduce service costs and can handle many customers simultaneously.”

As customer service chatbots are a service in themselves, they must be thought of and designed as such even by companies that do not have services at the center of their operations. According to Kimbell (2011), services can be approached from different stances, with marketing, operations and technology being some of them. There are several challenges to be observed in each of these fields. When AI-based services such as chatbots are popularized among companies and used daily by laypeople, ethical and human issues of artificial intelligence must also be observed.

The present work investigates whether or not customer service chatbots meet human-centric AI guidelines. The study’s main objective was to establish the relationship between these good practices and the increase or decrease of trust in the system and human autonomy senses. For that, the authors established which guidelines relate to each of these senses, explained possible applications of these approaches in the context of chatbots, and conducted a survey to check if users perceived those applications and whether meeting or not these propositions impacts these two feelings.



The need for explainable chatbot services

Corporations whose core business is the provision of services use AI improvements to make internal activities more efficient and offer improvements and customizations to their customers in record time. According to Morelli (2002, p.5), "(...) services come into existence at the same moment they are being provided and used" and, for those companies, being able to use AI in its best can make this overlapping of phases a great competitive advantage."

The development of intelligent conversational agents by many companies has both human and ethical dimensions as much-needed approaches to watch. In addition to happening in real-time, promoting that aforementioned competitive advantage, bots also enable direct interaction between humans and artificial intelligence.

Kim (2018) points out a profound and fundamentally structural problem: many companies see services as products. Consequently, interactions between individuals participating in a service are handled as if they are subject to quality control measures for the product instead of being approached more personally and humanized. The situation could worsen as services become more automated.

Regarding customer service chatbots, companies often see maintaining a standardized model as more cost-effective than hiring skilled employees and giving them the autonomy to conduct customer surveys and use feedback to improve the automated service. However, this approach may not provide the client with a satisfying experience, and the customer may become a detractor.

Delivering the best experience with chat-emulating bots should be a fundamental concern of every organization that offers them regarding the possible negative impacts of a frustrating interaction. Considering concerns related to respecting people's fundamental rights, the feelings of inability to deal with the machine or misunderstanding of its operation should not be feelings experienced by the user.

Many AI systems that are part of society's daily life do not have a level of interpretability regarding their functioning or their decisions accessible to lay users. The field of Explainable Artificial Intelligence (XAI), focused on ways to explain intelligent systems decisions to people, emerges from the need to clarify how these robots work for humans. (Danilevsky et al., 2020).

The XAI literature mentions the existence of two groups of users: technical users and lay users. Technical users would be researchers or specialists, and lay users would be end users who interact with or supervise automation but do not have the skills to unravel or understand system models. According to Jin and Youn. (2021), this last



segment, much more diversified and voluminous, is not the focus of research in the field so far.

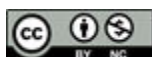
Even people with the knowledge to fit the definition of a technical user, when looking for chatbot-based customer assistance, have lay users' expectations: contact the company to resolve a practical situation. That said, this category of conversational agent that enables customer-brand contact when he/she needs help should be designed for lay users.

According to Rai's (2020) research, "(...) providing users with an effective explanation for the AI system's behavior can enhance their trust in the system." Similarly, the Google People + AI Guidebook (2019) defines trust as a willingness to take a risk based on the expectation of a benefit. When it comes to AI systems, users' trust in them can fluctuate based on their experiences using the system. Three key factors contribute to increasing trust: ability, reliability, and benevolence. Ability refers to a product's capability to meet users' needs, while reliability pertains to the consistency of this capability, ensuring that expectations are not broken. Benevolence refers to the belief that interactions with the AI system are mutually beneficial to both the user and the application.

The user has a problem that led to the conversation with the chatbot; if his/her goal of resolving his/her problem is not achieved, his/her trust in the product may be reduced or nullified abruptly. The machine's behavior, in this case, hurts the reliability and, consequently, the belief of competence of the system – two fundamental components in the construction of trust.

Clarity regarding AI models can also affect a range of emotions beyond just those related to trust. Sankaran and Markopoulos (2021) proposed a definition of human autonomy based on psychological principles, which highlights the importance of empowering individuals with the freedom to make their own choices and pursue their own goals. In the context of human-computer interaction (HCI), autonomy is often defined as the ability to make independent decisions and can be attributed to either the user or the system. However, when systems demonstrate a level of autonomy that exceeds the user's comprehension, it can lead to a sense of loss of control and ultimately undermine the user's sense of personal autonomy.

The black box nature of the algorithms involved in virtual attendants' decisions makes it difficult for users to understand and, therefore, can negatively impact this point of autonomy. If there is no clarification, the capacity for an independent decision with some objective already outlined, as stated in the definition of autonomy in HCI, may not be fully met.



Methodology

The guidebook mentioned above and Microsoft's Guideline for Human-AI Interaction are good practice guides that should be considered when developing any AI system, especially those with graphic user interfaces for laypeople. Based on experience, usability testing, and consumer research, each of these technology giants has defined guidelines that gather their know-how and their conclusions come together on several points.

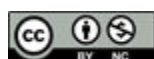
Both Google and Microsoft's recommendations combine statements that serve as guidelines for constructing and maintaining AI systems that consider the human being. Their publications emphasize that the development of AI-based systems must have the human being at the center of its process and accentuate the importance of transparency. This transparency includes clarifying decisions and promoting clear communication with the user, which would influence their feeling of trust in the technology and personal autonomy concerning the application.

Google's framework brings 23 recommendations – called patterns – to improve the explainability of the system and, consequently, the user's trust in it. Microsoft defined 18 topics with the same goals. These guidelines can also be related to the sense of human autonomy as defined earlier in the present study.

The two frameworks are equivalent. Some Google recommendations are analogous to sets of Microsoft advice – or, in a few cases, just one. For the present study, they were associated with definitions of trust or human autonomy. Subsequently, they were explained in just one statement each, and a way to meet it in the context of interaction with service chatbots was defined.

The definitions served as the basis for constructing a questionnaire whose answers would indicate whether the user understood that the topics were being attended or not. With closed and open questions - these not mandatory -, it is possible to assess the impact of the system's behavior on these two perceptions based on positive or negative experiences.

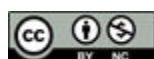
Table 1 shows the equivalence between Google and Microsoft guidelines and relates each group of topics of the guidebooks as influential in trust or autonomy.



Related sense	Google People + AI Guidebook	Microsoft AI Design Guidelines
Trust (T1)	Set the right expectations. Be transparent with your users about what your AI-powered product can and cannot do.	Make clear what the system can do. Help the user understand what the AI system is capable of doing. / Make clear how well the system can do what it can do. Help the user understand how often the AI system may make mistakes. / Make clear why the system did what it did. Enable the user to access an explanation of why the AI system behaved as it did.
Trust (T2)	Explain the benefit, not the technology. Help users understand your product's capabilities rather than what's under the hood.	Convey the consequences of user actions. Immediately update or convey how user actions will impact future behaviors of the AI system
Trust (T3)	Be accountable for errors. Understand the types of errors users might encounter and have a plan for resolving.	Support efficient correction. Make it easy to edit, refine, or recover when the AI system is wrong.
Trust (T4)	Explain for understanding, not completeness. Focus on giving your users the information they need in the moment, rather than a full run-down of your system.	Show contextually relevant information. Display information relevant to the user's current task and environment.
Autonomy (A1)	Make it safe to explore. Let users test drive the system with easily reversible actions.	Support efficient invocation. Make it easy to invoke or request the AI system's services when needed. / Support efficient dismissal. Make it easy to dismiss or ignore undesired AI system services.
Autonomy (A2)	Anchor on familiarity. As you onboard users to a new AI-driven product or feature, guide them with familiar touchpoints.	Match relevant social norms. Ensure the experience is delivered in a way that users would expect, given their social and cultural context.
Autonomy (A3)	Let users give feedback. Give users the opportunity for real-time teaching, feedback and error correction.	Encourage granular feedback. Enable the user to provide feedback indicating their preferences during regular interaction with the AI system.
Autonomy (A4)	Let users supervise automation. Maintaining control over automation helps users build comfort and correct when things go wrong.	Provide global controls. Allow the user to globally customize what the AI system monitors and how it behaves

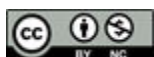
Table 1. Google's and Microsoft guidelines, their parity and the sense each one is related to

The analysis of each line on Table 1 allows a more direct example of what an intelligent conversational agent whose development considers the human factor



should do. Statements T1, T2, T3 and T4 correspond to trust and statements A1, A2, A3 and A4 are related to autonomy.

- (T1) States the need for clarification and the system's behavior forecasting. In human-chatbot interaction, it means making the user aware of that machine's limitations. The actual need for contact with a human being due to the robot's impossibility to solving specific problems must be evident along the interaction.
- (T2) States about ease of understanding system behaviors. It is necessary to clarify what action or user choice led to a particular path immediately so he/she can try again in a better way to reach his/her goal.
- (T3) States about communication efficiency and error correction. It means the system recognizes quickly that the user has identified a flaw or inconsistency and gives live and valuable feedback to help resolve the problem. The ease of returning to a previous chat stage is also here.
- (T4) States the need to enable plain communication avoiding cognitive overload. It means presenting helpful information in a specific interaction context. It is related to T2.
- (A1) States the need for an interaction simple enough to allow new tries and remaking the same path if wanted. It means making it easy to go back to previous steps or allowing new choices without having to restart the entire process.
- (A2) States the need to consider patterns that are familiar to the user. It means considering mental models already established in similar systems and adapting the experience.
- (A3) States the need to allow feedback and system customization. It means performing new queries as soon as the user communicates he/she couldn't achieve his/her goal. Collecting real-time feedback throughout the interaction is important, such as informing the user that it will be considered for the system's refinement.
- (A4) States the need for users to monitor system behavior and provide feedback. It means promoting user understanding about what each path will give him/her so he/she can ensure the system is performing as it should and assure the consideration of user's feedback for AI improvements.



Survey

To reach a wider audience quickly, a survey was utilized. This tool enables anonymity, which helps to decrease any inhibition participants may feel and encourages honest responses.

It had 18 questions – 3 demographic and 1 for selecting participants within the necessary population: users who interact or have interacted with service chatbots in their daily life. The other 14 questions were based on topics T1 to T4 and A1 to A4. Most inquiries were multiple choice with an evaluative or dichotomous nature – yes or no. Three questions had a 5-point scale, and two questions were open-ended.

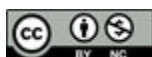
Each of the questions was planned for assessing trust, autonomy or both. They were written in Portuguese and in the topic of finding and discussion of the present study they will be presented translated and arranged next to the statements to which they are related – trust statements or autonomy statements.

Findings and Discussion

The first question, answered by 73 people, took respondents who did not have contact with chatbots in their daily lives to the end of the form. Other respondents were taken to the next question. Only 1 individual said he/she had no previous experience with chatbots, so the total number of respondents was 72.

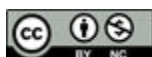


Figure 1. Demographic data showing most of the respondents were over 31 years old and graduated

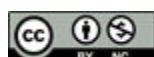


Demographic data shows over 90% of the respondent population is graduated or post-graduated and, therefore, has a good literacy level. This high reading proficiency minimizes the possibility of difficulties with erudite vocabulary – which is not usually present in service chatbots, but even if there were any of this particular case, most participants would understand the language. In addition, more than 60% of respondents are between 51 and 60 years old, making up a generation that has had extensive contact with other customer service arrangements.

Table 2 displays the questions asked to the respondents, the statements to which they relate, and the assessment of the question questioned concerning the ideas of trust in the system and human autonomy. This assessment indicates the prevalence of responses that point to a negative or positive perspective on each question. Open questions are not covered, only those that were dichotomous or had scales in which the extremes represented negative and positive perspectives.



Question	Related statements	Trust in the system	Human autonomy
Do you normally find the communication with a robot easy, whether by voice or texting?	T2, T4, A1	Negative	Negative
Has the virtual assistant ever provided you with far too many options while you were searching for more accurate information?	T2, T4, A3	Negative	Negative
Do the virtual assistants usually reply to you with appropriate information related to your search?	T2	Neutral to positive	Does not apply
If you answered differently from "always", can you tell the reasons to the assistants' error?	T2	Negative	Does not apply
Can you notify the assistant when it makes an error?	T2, T3, A1, A3	Negative	Negative
Do you feel in control during the interaction with a customer service chatbot?	T2, A2, A3, A4	Negative	Negative
Have you noticed you were able to guide the chatbot options towards the result you desire?	A3, A4	Negative	Negative
Do you notice it is possible to return to any preceding question if you want to change your previous answers?	A1	Does not apply	Negative
Some chatbots require a given number or spoken answer to move on to the next step. Can you normally predict the next step after you key in or say the answer you desire?	T1, T2, A2	Neutral to positive	Neutral to positive
Can you notice any standardized voice and/or message from the robots?	A2	Does not apply	Positive



Question	Related statements	Trust in the system	Human autonomy
Do you think there is any significant difference in the robot assistance from different companies?	A2	Does not apply	Positive
Have you tried contacting companies to suggest any improvements in their chatbots?	A3	Does not apply	Negative

Table 2. Single questions, their related statements and respondent's perspective when answering each question

Regarding expectations alignment (T1), participants were asked about their level of awareness about the probable next steps when answering a chatbot question. Almost 40% of the respondents were neutral, choosing the central option of the scale. More users, however, said they were aware of the following steps (25 people) compared to those who declared they never or rarely knew what they would be exposed to from then on (17 people). These answers also relate to the trust guideline for ease of understanding (T2).

Questions about familiarity with the systems or recurrence of already known patterns (A2) revealed a perception of similarity between different customer service robots. 60 of the 72 participants said they noticed patterns; only 25 said they saw significant differences between robots from different brands. This shows a tendency for companies to reproduce in their chatbot behaviors already known from previously existing automation. However, being similar does not mean they are human-centered AI; there is no indication that this standard is adopted for this purpose and not just because it is easily replicable. In any case, the fact that it is familiar to the user is favorable for the feeling of autonomy.

Some answers to open questions also brought information related to topics T2 and A1. On the subject of repeating systems patterns, one interviewee said "most chatbots work as if they were an ARU". ARU is an Audible Response Unit, a class of automated services popular since the 1990s. Regarding the differences between chatbots according to the responsible organization, another respondent said "Narturgy's is terrible. Itaú's is great. There is no rule, it depends on who invests seriously and enough in this modality." The statements show that a standard familiar to the user is followed, but he/she believes, from previous experiences, that some particular companies finance this service improvement. It is noticeable that a good



and a bad experience with different chatbots are still in that consumer's memory and he/she associates it with the brand.

6 questions mentioned the ease of understanding system's behaviour or possible paths (T1, T2, A4). Only 13 of the 72 respondents stated they rarely receive information adequate to their request, and most remain neutral (36.8%) or do not understand the reasons that lead to missteps (33.9%). Some statements regarding the service bringing or not the requested answer were "The virtual attendant shows standard responses, which normally do not apply to the case of those who seek service.", "Often, the option I need does not appear, and it becomes a waste of time.", "Often the options do not meet what we want to report" and "The very concise answer of the robot does not always suit me."

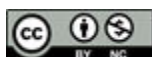
The ease of returning to previous steps (A1) along with the clarity of its limitations (T1) was mentioned in a written response: "I get stuck in predetermined algorithms and waste time talking to virtual attendants. They cannot flow the service". Regarding system customization based on feedback (A3) and user supervision (A4), it was said that "You often get into a loop. And bots have a low ability to understand, and of course, you cannot put cordialities (good morning)". The question about whether or not to drive a chatbot through a trail they would like brought 39 negative responses among the 72 participants.

Approximately 61% of respondents said interacting and communicating with the chatbot is not easy, which is a perception associated with uncomplicated interaction (A1) and monitoring system's behavior (A4). Additionally, around 88% of people said they do not feel in control when interacting with the bot while only 2% said they have this sense of autonomy.

Conclusions

Failure to comply with human-centered AI best practices guidelines by applications classified as chatbots is, from the responses obtained, a factor that negatively impacts trust and the feeling of human autonomy.

Negative experiences with virtual agents are constantly shared by customers, who often mention discontentment with conversational agents. However, the organizations do not observe and work on the possible consequences of this disappointment. The present study shows, in a preliminary way, that usual human-chatbot interaction must be developed not only with a technological approach; user experience, service design, and artificial intelligence ethics have to be taken into



consideration in order to promote a better acceptance of this type of service by consumers and generate better sensations.

In the population studied, these virtual assistants do not reflect the statements placed as a basis for a good experience of the human being with a system based on artificial intelligence. As a participant reported, "Many companies don't put the user at the center of the solution. In this way, the chatbot becomes an obstacle in the delivery of the service."

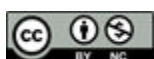
The issue of reliability, which influences the establishment of trust, is directly impacted in the moments of frustration raised by the respondents. The limitation observed in the system's efficiency when compared to customer assistance service performed by a person and the recurrence of negative terms such as "irritation" and "hate" in responses about solving their problems reinforces this finding. The absence of options that meet their needs in the systems, straightforward in open responses, and the perception that the virtual attendant represents an obstacle and not a facilitator hurts the pillar of benevolence - also placed in the definition of trust.

The opacity of the AI chatbots also hurts the sense of personal autonomy placed as a basis for a pleasing interaction. There is a lack of clarity about decisions made by the robot and how to provide them feedback on errors. The users mentioned they do not notice improvements on the chatbots, suggested by them or not, even if staying a long time without using the same service.

Limitations and Suggestions for Future Research

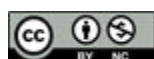
Some points not covered in the scope of this study, which addressed the relationship with customer service chatbots, can be of great value. Cutouts considering chatbots from companies in specific sectors will be of great value for business and customer relationship fields. Evaluating conversational agents provided by public institutions and recruiting respondents with a lower literacy level is also an alternative way to extract significant value for public policy development. People who use public health services or have children who go to public education institutions, for example, could gain tremendous support with a focused improvement of chatbots in these sectors.

Finally, it is possible to carry out in future research a deepening from cultural differences in the perceptions of trust and autonomy, as well as possible disparities between generations. User sense of trust and autonomy considering chatbots of different structures, such as voice and message, or designed for different channels, such as WhatsApp or the organization's institutional website, can also be compared. Hopefully, it will be possible to improve general guidelines in order to clarify how to apply them to each type of AI-based system.



References

- Adamopoulou, E., Moussiades, L. (2020) *Chatbots: History, technology, and applications*. Published in *Machine Learning with Applications*, V. 2. Retrieved from <https://www.sciencedirect.com/science/article/pii/S2666827020300062>
- Danilevsky, M. Qian, K. Aharonov, R. Katsis, Y. Kawas, B. and Sen, P (2020). *A Survey of the State of Explainable AI for Natural Language Processing*. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 447–459, Suzhou, China. Retrieved from <https://aclanthology.org/2020.aacl-main.46/>
- Google. *People + AI Guidebook Patterns*. Available on <https://pair.withgoogle.com/guidebook/patterns>
- Jin, V., Youn, S. (2021) In “A.I. we trust?” The effects of parasocial interaction and technopian versus luddite ideological views on chatbot-based customer relationship management in the emerging “feeling economy”. Published in *Computers in Human Behavior* V. 119. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0747563221000431>
- Kimbell, L. (2011) *Designing for Service as One Way of Designing Services*. Retrieved from https://www.researchgate.net/publication/282989518_Designing_for_Service_as_One_Way_of_Designing_Services
- Microsoft. *Guidelines for Human-AI Interaction*. (2019). Published in CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. Retrieved from <https://www.microsoft.com/en-us/research/uploads/prod/2019/01/Guidelines-for-Human-AI-Interaction-camera-ready.pdf>
- Kim, M. (2018) *Designing for Participation: Dignity and Autonomy of Service (Part 2)*. *Design Issues* 2018; 34 (3): 89–102 Retrieved from <https://direct.mit.edu/desi/article/34/3/89/69323/Designing-for-Participation-Dignity-and-Autonomy>



- Morelli, N. (2002) *Designing Product/Service Systems: A Methodological Exploration*. Published in Massachusetts Institute of Technology Design Issues: Volume 18, Number 3 Retrieved from <https://www.cs.cmu.edu/~jhm/Readings/Morelli.pdf>
- Nicolescu, L., Tudorache, M.T. (2022) *Human-Computer Interaction in Customer Service: The Experience with AI Chatbots—A Systematic Literature Review*. Retrieved from <https://www.sciencedirect.com/science/article/pii/S2666827020300062>
- Rai, Arun. (2020, April). *Explainable AI: from black box to glass box*. *J. of the Acad. Mark. Sci.* (p. 137-141). Retrieved from <https://link.springer.com/article/10.1007/s11747-019-00710-5>
- Reunanen, N., Falay, Z., Roto, V., Vaajakallio, K. (2021, February). *Combining machine learning and Service Design to improve customer experience*. Paper published in ServDes.2020 Tensions, Paradoxes, Plurality: 2021 Conference Proceedings (pp. 124-133). Retrieved from https://acris.aalto.fi/ws/portalfiles/portal/78737352/Reunanen_FvF_Roto_Vaajakallio_Combining_machine_learning_and_service_design.pdf
- Sankaran, S., Markopoulos, P. (2021, June). *"It's like a puppet master": User Perceptions of Personal Autonomy when Interacting with Intelligent Technologies*. Paper published in Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization UMAP'21 (p.108–118). Retrieved from <https://dl.acm.org/doi/10.1145/3450613.3456820>

