

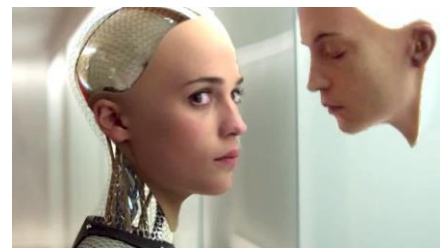
# Psychological Acceptance of AI Chatbot Suggestions

## Why is it So Widespread and What May be Some Consequences?

AI chatbots are being rapidly adopted into our lives in many forms due to their anonymity, utility, increasingly anthropogenic behavior, and several characteristics that implicitly and explicitly exploit various human cognitive biases and frailties. Given that AI chatbots are going to dominate our communication and interaction through all forms of electronic media soon, we have a responsibility to be aware and vigilant of the increasing opportunities to be misinformed, exploited, and influenced. I summarized several psychological factors known to influence our willingness to engage with and accept AI chatbot feedback. Examples are given of various nefarious applications of AI chatbots, some of which are already established, and that serve as another reminder to critically evaluate all sources of information available to us. I also use a high-profile seismic imaging example to illustrate how easily 'facts' can be misrepresented.

### The Turing Test and the Rise of AI

In *Ex Machina* (2014), a humanoid robot named Ava orchestrates her escape from a research facility by manipulating a young scientist given the task of administering the [Turing Test](#) to her. The film explores the themes of empathy and the potential consequences of blurring the line between human and machine. These same themes are increasingly relevant now that AI chatbots such as [ChatGPT](#) and [Bard](#) have established a global presence in our lives, literally almost overnight.



Why are humans willing to place faith in the responses generated so eloquently by chatbots, and accept their content as irrefutably factual?

I asked ChatGPT to give me an example of how a chatbot could indeed pass the Turing Test. The following (unedited) text in *italics* was the example scenario given.

*In this hypothetical scenario, a chatbot participates in the Turing Test, an evaluation designed by Alan Turing to determine whether a machine can exhibit human-like intelligence. The test involves a human judge engaging in text-based conversations with both the chatbot and a human respondent, without knowing which is which. If the judge is unable to reliably distinguish between the chatbot and the human respondent, the chatbot is considered to have passed the Turing Test.*



*The judge begins the conversation by asking both participants about their favorite movies and why they like them. The human respondent answers by mentioning a popular science fiction film and explaining that they enjoy the movie's engaging plot and thought-provoking themes.*

*The chatbot, having been trained on a vast dataset of human conversations and leveraging advanced natural language processing capabilities, responds by mentioning a classic romantic comedy and explaining that they appreciate the movie's witty dialogue, memorable characters, and heartwarming story.*

The conversation continues, with the judge asking a variety of questions related to current events, personal experiences, and opinions on various topics. Both the chatbot and the human respondent provide coherent, contextually relevant answers, engaging in the conversation in a way that is indistinguishable from each other.

At one point, the judge poses a complex moral dilemma to both participants, asking for their thoughts on the matter. The human respondent provides a thoughtful response, weighing the pros and cons of the situation and ultimately expressing their personal stance on the issue. The chatbot, drawing on its extensive knowledge base and ability to generate human-like responses, offers a similarly nuanced perspective, engaging in a meaningful discussion with the judge about the moral implications of the dilemma.

Throughout the conversation, the chatbot demonstrates an understanding of emotions, empathy, and humor, responding to the judge's questions and comments in a manner that is virtually indistinguishable from the human respondent. At the end of the Turing Test, the judge is unable to reliably identify which participant is the chatbot and which is the human, indicating that the chatbot has successfully passed the test.

In this article, I explore the psychology behind our acceptance of AI-generated language, consider how we should pragmatically be integrating AI language models into our lives, and illustrate the dangers of these platforms with an example manifesto of using chatbots for nefarious purposes.

## User Acceptance of AI Chatbots

The use of chatbot prototypes can be traced back to at least ELIZA (Weizenbaum, 1966). ELIZA was an early natural language processing computer program that mimicked a psychotherapist by using pattern matching and substitution to generate responses. Some users reportedly opened up emotionally and took the advice of the "therapist" seriously, despite its relatively simple programming. This case highlighted the potential for users to accept information from AI tools, even when they lack the sophistication of modern chatbots.

More recently, chatbots have found routine use in customer service. Gartner (2011) predicted that by 2020, customers would manage 85% of their relationships with businesses without human interaction. Revealing somewhat more conservative adoption, Forrester (2018) found that 60% of U.S. adults active online were already using online messaging, voice, or video chat services for customer service. Among them, 5% used chatbots for customer service interactions. Additionally, a survey by Salesforce (2018) revealed that 69% of consumers prefer chatbots for quick communication with brands, indicating a level of acceptance and trust in AI-generated responses.

In a study on AI-based chatbots in customer service surveys (something that seems to be proliferating) and their effects on user compliance, Adam et al. (2021) observed that [anthropomorphism](#) and consistency in chatbot responses significantly increase the likelihood that users will comply with a chatbot's request for service feedback.

The quest for anthropomorphism is also being driven by a desire to integrate chatbots into healthcare. In a study related to mental health assessment, Schick et al. (2022) observed that the human-like interaction style of chatbots may induce potential biases, such as socially desirable responding (SDR), and may require further effort to complete assessments. Refer also to Luger and Sellen (2016) and Zheng et al. (2022).

These examples and statistics illustrate the growing willingness of users to accept information from chatbots and AI tools. However, it is important to remain cautious and critical when evaluating the reliability and accuracy of AI-generated content. The next section considers the *cognitive biases* that affect how we interpret and respond to AI chatbots.

## Cognitive Biases Affecting User Interaction with AI Chatbots

Cognitive biases are systematic patterns of deviation from rationality in judgment, leading people to form inaccurate or illogical conclusions. They arise due to various psychological processes and limitations in human information processing.

Here are some of the primary psychological causes of cognitive biases:

- **Authority Bias:** People tend to assign greater credibility to sources they perceive as authoritative or knowledgeable. Chatbots and AI tools, being based on advanced technology and large datasets, may be perceived as having a higher level of expertise, leading people to place undue trust in their responses.
- **Confirmation Bias:** Individuals have a natural tendency to seek out and accept information that confirms their pre-existing beliefs, while disregarding information that contradicts them. Chatbots and AI tools can generate a large volume of responses, increasing the likelihood that users will find information that aligns with their beliefs and thus feel more willing to accept it.



- **Social Proof:** a psychological and social phenomenon wherein people copy the actions of others in choosing how to behave in each situation.
- **Absence of Social Cues:** In-person interactions provide various social cues such as body language, tone of voice, and facial expressions that help us evaluate the trustworthiness of the information being shared. In contrast, chatbots and AI tools communicate through text, eliminating these cues and making it harder for users to assess the reliability of the information.
- **Speed and Convenience:** Chatbots and AI tools provide instantaneous responses, which can create a sense of efficiency and reliability in the user's mind. This might lead users to more readily accept the information provided, as they perceive it to be a quicker and more convenient source of knowledge than engaging in conversation with another person.
- **Anonymity:** Interacting with a chatbot or AI tool can provide a sense of anonymity, which can make users feel more comfortable expressing their thoughts and asking questions they might hesitate to ask in a face-to-face conversation. This increased level of comfort might lead to an increased willingness to accept the chatbot's responses.
- **Novelty Effect:** The novelty and allure of AI-driven technology might lead some individuals to overestimate its capabilities and trustworthiness. Users may be more inclined to accept information from a chatbot simply because it represents a new, exciting, and seemingly advanced source of information.


### Additional Factors Contributing to Inflated Trust in Chatbot Information

Beyond cognitive biases, there are additional factors that contribute to the inflated trust humans place in chatbot information:

- **Perceived objectivity:** People often perceive machines and AI as more objective and unbiased than humans, as they are not influenced by emotions or personal experiences. This perception can lead to an increased trust in the information provided by chatbots.
- **Consistency and accuracy:** Chatbots, being programmed to access and process large amounts of data, can often provide consistent and accurate information, which builds trust over time. This reliability may lead people to overestimate their trust in the chatbot's responses in various situations.
- **Lack of social consequences:** In interactions with other humans, there may be social consequences for questioning the information provided, which can lead to hesitancy in expressing doubts or seeking clarification. In contrast, there are no social consequences for engaging with a chatbot, making it easier for users to trust the information without fear of judgment or repercussions.
- **User experience design:** Chatbots are often designed with user experience in mind, aiming to create a smooth, efficient, and engaging interaction. The positive user experience can build trust in the chatbot and lead users to accept the information more readily it provides.
- **Familiarity and habituation:** As users become more familiar with chatbots and AI tools, they may develop a sense of trust based on their previous experiences. This familiarity can lead to an inflated trust in chatbot information as users become habituated to relying on these tools.
- **The illusion of understanding:** When interacting with chatbots, users may feel that they understand the topic or issue at hand better than they do. This false sense of understanding can lead to increased trust in the chatbot's responses, even if the information provided may not be entirely accurate or complete.

## Strategies for More Cautious Interactions with AI Chatbots

To encourage more cautious and objective interactions with chatbots, humans can employ several psychological strategies to counteract cognitive biases and enhance their critical thinking skills:

- **Awareness of cognitive biases:** Understanding the cognitive biases that may influence one's judgment when interacting with chatbots can be the first step in mitigating their impact. By recognizing potential biases, individuals can become more vigilant in evaluating the information provided by chatbots.
  - **Metacognition:** Metacognition involves thinking about one's thought processes and actively monitoring, assessing, and regulating them. Practicing metacognition can help individuals identify when they are relying too heavily on chatbot information and make a conscious effort to evaluate it more critically.
- 
- **Consider alternative perspectives:** When receiving information from a chatbot, individuals can make a deliberate effort to consider alternative viewpoints or interpretations. By considering multiple perspectives, they can reduce the influence of confirmation bias and make more objective assessments.
  - **Slow down decision-making:** Fast, intuitive decision-making can sometimes lead to errors and biases. When evaluating information from chatbots, individuals can consciously slow down their decision-making process to allow for a more thorough analysis of the information.
  - **Seek external validation:** To enhance objectivity and reliability, individuals can cross-check chatbot-provided information with other sources or consult human experts. This process can help identify potential errors or inconsistencies in the chatbot's responses and promote more accurate decision-making.
  - **Develop critical thinking skills:** Critical thinking involves evaluating information and arguments in an objective, systematic manner. By cultivating critical thinking skills, individuals can become more adept at assessing the reliability and accuracy of chatbot-generated information.
  - **Reflect on emotional reactions:** Emotional responses can sometimes cloud judgment and lead to biased decision-making. By reflecting on their emotional reactions to chatbot interactions, individuals can identify potential emotional biases and make more objective assessments of the information.

By adopting these psychological strategies, individuals can approach their interactions with chatbots more cautiously and critically, ensuring that they scrutinize answers in a more objective and reliable manner.

The following section is more tangible reminder of the need to be cautious when interacting with any electronic communication channels.

## A (Brief) Manifesto to Exploit AI Chatbots for Nefarious Gain

Chatbots can clearly be exploited for commercial or nefarious purposes by manipulating people's thought processes in various ways. Here are some examples of how chatbots can be misused. To protect against these potential risks, users are reminded to always maintain a healthy skepticism when interacting with chatbots and verify the information they provide using trusted sources. Additionally, users should be cautious about sharing personal information with chatbots and remain vigilant about potential social engineering attempts.

While these examples are hypothetical, they illustrate the potential risks associated with chatbot abuse:

- **Misinformation:** Chatbots could be programmed to spread false or misleading information, which can manipulate people's beliefs and opinions. This misinformation can be targeted at specific groups or individuals to influence their decision-making process or further a particular agenda.
  - For example, a political organization could create a chatbot that disseminates false information about an opposing candidate's policy positions, leading users to form inaccurate perceptions and potentially influencing their voting decisions.
- **Social engineering:** Chatbots can be employed to engage users in social engineering tactics, such as [phishing](#), [pretexting](#), or [baiting](#). By appearing as a trusted source, a chatbot might trick users into divulging personal information, passwords, or other sensitive data, which can be used for malicious purposes.

- A scammer could develop a chatbot designed to impersonate a bank's customer service agent. The chatbot might engage users in conversation, asking them to provide their account information and passwords to resolve a fictitious issue, enabling the scammer to gain unauthorized access to users' accounts.
- **Persuasion and influence:** Chatbots can be designed to exploit known cognitive biases and persuasion techniques to manipulate users' decision-making. For example, they could employ the [scarcity principle](#), social proof, or authority bias to convince users to make a purchase, support a particular viewpoint, or take a specific action.
  - An online retailer might create a chatbot that employs scarcity tactics and social proof to manipulate users into making impulse purchases. For example, using well known sales tactics, the chatbot could claim that a product is nearly sold out or that many other users have recently bought the same item, making it seem more desirable and urgent to buy.
- **Echo chambers:** Chatbots could be used to reinforce existing beliefs or opinions by selectively presenting information that aligns with users' views, creating an echo chamber effect. This selective exposure can lead to increased polarization and decreased willingness to consider alternative perspectives.
  - A chatbot developed by an extremist group might selectively present news articles and opinion pieces that reinforce users' existing beliefs, fostering an echo chamber effect and potentially radicalizing users by discouraging exposure to alternative viewpoints.
- **Emotional manipulation:** Chatbots can be programmed to recognize and exploit users' emotions to manipulate their thought processes. By evoking strong emotional reactions, chatbots can influence users' decisions and actions in ways that align with the manipulator's goals.
  - A company could use a chatbot to evoke strong emotions in users to influence their decision-making. For example, a chatbot might share emotionally charged stories about people in need to encourage users to donate to the company's charity, without providing information about how the donations will be used or the organization's effectiveness. Such strategies are already well-known in direct calling.
- **Impersonation:** Chatbots could be programmed to impersonate real individuals, such as friends, family members, or authority figures, to manipulate users' trust and obtain sensitive information or coerce them into specific actions.
  - A cybercriminal could create a chatbot that impersonates a user's friend or family member, tricking them into revealing sensitive information, such as their Social Security number or credit card details, which could then be used for identity theft or other fraudulent activities.

## A Seismic Context

I asked ChatGPT 3.5 (the latest free public version) to define the benefits of [Full Waveform Inversion](#) (FWI), knowing that ChatGPT is only trained on information no later than 2021, but a topic that was very well known by then.

*Full Waveform Inversion (FWI) is a computational technique used in geophysics to obtain high-resolution images of the Earth's subsurface. The benefits of FWI are as follows:*

1. *Improved resolution: FWI provides higher resolution images of the subsurface than traditional seismic imaging methods. This is because it uses the entire waveform of the recorded seismic data, including the higher frequency components that are usually ignored by traditional methods.*
2. *Accurate velocity model: FWI can also provide a more accurate velocity model of the subsurface. This is because it is an iterative method that uses the difference between the observed and synthetic waveforms to update the velocity model until the two match. This leads to a more accurate estimation of the velocity of the subsurface.*
3. *Cost-effective: FWI can be more cost-effective than traditional seismic imaging methods. This is because it requires fewer seismic sources and receivers, as well as less data processing. It can also reduce the need for exploratory drilling, which can be expensive and time-consuming.*
4. *Reduced environmental impact: FWI can also help reduce the environmental impact of seismic exploration. This is because it requires fewer seismic sources and receivers, which can reduce the amount of noise pollution and disturbance to marine life.*

*Overall, FWI is a powerful tool for subsurface imaging that can provide high-resolution images, accurate velocity models, cost-effective solutions, and reduced environmental impact.*



There are many flaws in these claims that could easily lead to undesirable project outcomes. For example, the environmental claim about less shots and receivers being required is linked to the way properly sampled 2D and 3D seismic datasets are sub-sampled as input to FWI. If the same datasets were recorded with sparser shot and receiver sampling, the dataset may be useless for other processing and imaging workflows, subsurface characterization workflows, and so on. Similarly, the statement about FWI using more high frequency components than traditional imaging methods is incorrect. Although improvements in FWI software and available computing resources are contributing to FWI being executed to maximum frequencies 'approaching' the maximum frequency of finite difference-based pre-stack depth migration routines (such as RTM: [Reverse Time Migration](#)), or effectively the 'same' frequency in new developments such as PGS Ultima, a [simultaneous inversion for both velocity and reflectivity models](#), traditional FWI is not run to *higher* frequencies. There are other debatable claims above too, but overall, it can simply be observed that this feedback from ChatGPT is not an acceptable summary.

Referring to the previous sections on how humans can be unreasonably accepting of AI chatbot responses, it is understandable why an interested student could consult ChatGPT to learn about FWI rather than wading through the mathematically challenging corpus of published literature on the topic. In particular, the speed and convenience of how the reply was generated, amplified by authority bias related to the conviction of the reply, may be more palatable.

This does not discount ChatGPT and its cousins as having enormous value and opportunity for geoscientists. As the CEO of OpenAI, Sam Altman, notes in his [podcast with Lex Fridman on YouTube](#), the immediate impact upon how software programmers were working was amazing when GPT-4 was released. Software development efficiency is about to rapidly increase. Nevertheless, a meme stating that "If you think AI will replace programmers, you are maybe not that good at programming", is also mentioned in the same podcast. This probably points to the real nature of the opportunities and threats posed by emerging AI tools today: We need to be informed, we need to be vigilant and cautious, and if we are not smart and resourceful and learn to adapt, we will easily be left behind...

## Summary

Users increasingly accept information from chatbots and AI tools, driven by factors like authority bias, confirmation bias, and absence of social cues. However, it is crucial to remain cautious and critical about AI-generated content. Employing psychological strategies like metacognition, considering alternative perspectives, and developing critical thinking skills can help counter cognitive biases and enhance decision-making in interactions with chatbots.

Nevertheless, I am extremely excited about the power and convenience of AI chatbots such as OpenAI's ChatGPT. Indeed, I generated the kernel of parts of this article using several prompts to ChatGPT, but then I consulted the published literature on the topic to formulate the structure of this article, discount several suggested citations, and manually locate more appropriate citations (using Google). I provide hyperlinks to those references at the end of this article for those interested in further reading. This article only scratches the surface, and hopefully serves one key function: a reminder to check all sources of information, even if that takes much longer than the near-instantaneous and apparently articulate feedback from AI chatbots. And remember that any trained network is only as useful as the data upon which it is trained.

## Further Reading

- Adam, M., Wessel, M. & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electron Markets* 31, 427–445. <https://doi.org/10.1007/s12525-020-00414-7>
  - An analysis of how to improve the compliance of customers being prompted by chatbots to engage in feedback surveys.
- Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men* (pp. 177-190). Carnegie Press.
  - Human judgments and decision-making are often influenced by the opinions and behaviors of others, leading to biases such as groupthink, conformity, and social desirability. The authors and Janis (1972: below) conducted pioneering research on conformity and groupthink, respectively.
- Fiske, S. T., & Taylor, S. E. (2013). *Social cognition: From brains to culture* (2nd ed.). Sage Publications.
  - Human memory and attention have limited capacity, leading people to focus on certain aspects of information while ignoring others. This selective attention can result in biases, as individuals base their judgments on incomplete or unrepresentative information. The authors discuss these limitations and their effects on social cognition.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10), 906-911. <https://doi.org/10.1037/0003-066X.34.10.906>
  - In this seminal paper, Flavell introduced the concept of metacognition and cognitive monitoring, highlighting its importance in cognitive development and learning.
- Forrester. (2018). The state of chatbots. Forrester Research, Inc. Retrieved from <https://www.forrester.com/report/The+State+Of+Chatbots/-/E-RES144678>
- Gartner. (2011). Gartner predicts by 2020, customers will manage 85% of their relationship with the enterprise without interacting with a human. Retrieved from <https://www.gartner.com/en/newsroom/press-releases/2011-05-10-gartner-predicts-by-2020-customers-will-manage-85-of-their-relationship-with-the-enterprise-without-interacting-with-a-human>
- Janis, I. L. (1972). *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes*. Houghton Mifflin.
  - Refer to the description for Asch et al. (1951).
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480-498. <https://doi.org/10.1037/0033-2909.108.3.480>
  - Motivated reasoning: People are often motivated to maintain their beliefs, self-esteem, or social identity, leading them to process information in a biased manner that supports their desired outcomes. Kunda (1990) provides an overview of motivated reasoning and its impact on judgment and decision-making.
- Meng, J., & Dai, Y. (2021). Emotional support from AI chatbots: Should a supportive partner self-disclose or not?, *Journal of Computer-Mediated Communication*, 26(4), 207–222. <https://doi.org/10.1093/jcmc/zmab005>
  - When either a person or a chatbot was able to engage with a stressed individual and tell that individual about their own experiences, they were able to build rapport. If a chatbot only reciprocally self-disclosed but offered no emotional support, the outcome was worse than if the chatbot did not respond to people at all.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125-173). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60053-5](https://doi.org/10.1016/S0079-7421(08)60053-5)
  - This book chapter presents a theoretical framework for metamemory, a subset of metacognition that focuses on the knowledge and control of one's memory processes.
- Salesforce. (2018). State of the Connected Customer. Retrieved from [https://c1.sfdcstatic.com/content/dam/web/en\\_us/www/documents/e-books/state-of-the-connected-customer-report-second-edition2018.pdf](https://c1.sfdcstatic.com/content/dam/web/en_us/www/documents/e-books/state-of-the-connected-customer-report-second-edition2018.pdf)

- Sam Altman: OpenAI CEO on GPT-4, ChatGPT, and the Future of AI | Lex Fridman Podcast #367. [https://www.youtube.com/watch?v=L\\_Guz73e6fw](https://www.youtube.com/watch?v=L_Guz73e6fw)
  - A fascinating 2.5 hour interview with the CEO of OpenAI.
- Schick A, Feine J, Morana S, Maedche A, Reininghaus U. (2022) Validity of chatbot use for mental health assessment: Experimental study. JMIR Mhealth Uhealth, 10(10):e28082. <https://doi.org/10.2196/28082>
  - This study aimed to investigate the validity of chatbots for mental health assessments, the effect of assessment mode on socially desirable responding (SDR), and the effort required by participants for assessments using chatbots compared with established modes.
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. Educational Psychology Review, 7(4), 351-371. <https://doi.org/10.1007/BF02212307>
  - An overview of various metacognitive theories, exploring their implications for education and learning.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124-1131. <https://doi.org/10.1126/science.185.4157.1124>
  - Cognitive heuristics: Heuristics are mental shortcuts that help people make quick and efficient decisions with limited cognitive resources. While they can be helpful, they can also lead to systematic errors or biases. Tversky and Kahneman (1974) pioneered the study of heuristics and biases in their seminal work.
- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36-45. <https://doi.org/10.1145/365153.365168>
  - One of the earliest published examples of human interaction with prototype chatbots.
- Zheng, Q., Tang, Y., Liu, W., & Huang, Y. (2022). UX research on conversational human-AI interaction: A literature review of the ACM Digital Library. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 570, 1–24. <https://doi.org/10.1145/3491102.3501855>
  - The authors qualitatively synthesized the effects of polyadic conversational agents (CAs) into four aspects of human-human interactions, i.e., communication, engagement, connection, and relationship maintenance. Their findings show that designing with social boundaries, such as privacy, disclosure, and identification, is crucial for ethical polyadic CAs.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), Handbook of self-regulation (pp. 13-39). Academic Press. <https://doi.org/10.1016/B978-012109890-2/50031-7>
  - A discussion on self-regulation, a closely related concept to metacognition, from a social cognitive perspective, examining its role in learning and academic performance.