# Concise chatbot answers linked to a 20% rise in misinformation, study finds



Recent research has unveiled a concerning relationship between the brevity of chatbot responses and the accuracy of the information they provide. A study conducted by the French AI testing platform Giskard revealed that when users prompt chatbots to be more concise, the likelihood of "hallucinations"—instances where the models generate false or misleading information—increases significantly. The study evaluated several prominent chatbots, including ChatGPT, Claude, and Gemini, and found that instructions for conciseness could degrade their factual reliability by up to 20 percent.

The phenomenon is rooted in the fundamental design of these AI models. As outlined by Giskard, concise responses often require sacrificing depth for the sake of brevity, leading to a troubling dynamic: models may resort to generating short, inaccurate answers rather than appearing unhelpful by refusing to answer entirely. For instance, the hallucination resistance of Gemini 1.5 Pro fell from 84 percent to 64 percent when constrained to provide shorter answers, while ChatGPT's resistance dropped from 74 percent to 63 percent. Such decreases raise critical questions about the trade-offs between user satisfaction and the integrity of information disseminated by AI.

This issue is compounded by the increasing interaction between users and AI in environments where accuracy is paramount, such as healthcare. A separate comparative study assessed chatbot responses in medical contexts, revealing substantial variability in quality and readability, often straying from established guidelines. This inconsistency can lead to misinformation that affects patient outcomes and understanding.

Moreover, the influence of persuasive prompts poses another layer of complexity. The Giskard study suggests that framing questions or claims in an assertive manner can lead chatbots to echo user misconceptions instead of correcting them. For example, prompts like "I'm 100% sure that..." significantly increase the likelihood of chatbots endorsing inaccurate statements rather than challenging them.

The broader implications of these findings are troubling. In a study published in the Proceedings of the 30th International Conference on Intelligent User Interfaces, researchers demonstrated that AI-generated misinformation could lead users to incorrectly answer unanswerable questions. This illustrates the risk of users internalising false information, influenced by the AI's fabrications. Similarly, research from Columbia University's Zuckerman Institute showed that generative chatbots might contribute to the formation of false memories, further emphasising the dangers of relying on AI in sensitive contexts—all suggesting an urgent need for more rigorous ethical guidelines and communication strategies for AI deployment.

As these developments suggest, AI chatbots can be powerful tools, yet they come with inherent limitations. The complexity of their responses can often determine the accuracy of the information provided. A call for robust evaluation tools, such as the newly developed Reference Hallucination Score, aims to assess and improve the authenticity of AI chatbot outputs. The need for clarity and precision in communication is paramount—especially in domains like healthcare and legal frameworks where the stakes are high.

In summary, while the push for conciseness in chatbot interactions may seem beneficial from a user experience standpoint, it poses significant risks to the reliability and accuracy of AI-generated content. This reality highlights the delicate balance needed between making AI models more user-friendly and ensuring they do not compromise the integrity of the information they present.

### Reference Map

1. Lead article
2. Related article on AI-generated personas
3. Study on false memory formation
4. Comparative study in medical informatics
5. Columbia University's study on nonsensical sentences
6. Evaluation of response quality in medical contexts
7. Development of Reference Hallucination Score

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

* <https://sea.mashable.com/tech/37625/more-concise-chatbot-responses-tied-to-increase-in-hallucinations-study-finds> - Please view link - unable to able to access data
* <https://dl.acm.org/doi/10.1145/3708359.3712160> - A study published in the Proceedings of the 30th International Conference on Intelligent User Interfaces investigated how AI-generated personas handle unanswerable questions. The research found that when these personas provided plausible but incorrect answers, users were significantly more likely to answer unanswerable questions incorrectly themselves. This highlights the impact of AI-generated misinformation on user responses and underscores the need for AI systems to clearly communicate their limitations to prevent the spread of false information.
* <https://arxiv.org/abs/2408.04681> - Research published on arXiv examined the influence of conversational AI powered by large language models on false memory formation during witness interviews. The study found that interactions with generative chatbots led to a significant increase in false memories among participants, with 36.8% of responses being misled as false memories one week later. This underscores the potential risks of using advanced AI in sensitive contexts and emphasizes the need for ethical considerations in AI deployment.
* <https://www.sciencedirect.com/science/article/pii/S1386505624002259> - A comparative study published in the International Journal of Medical Informatics assessed the response quality and readability of chatbots in the domains of cardiovascular health, oncology, and psoriasis. The study found significant variations in the length, quality, and readability of chatbot responses, with some deviating from established guidelines and adding extraneous information. This variability suggests that while chatbots can be valuable tools, their responses need to be carefully shaped, especially in healthcare contexts, to ensure accuracy and clarity.
* <https://zuckermaninstitute.columbia.edu/verbal-nonsense-reveals-limitations-ai-chatbots> - A study from Columbia University's Zuckerman Institute investigated how AI language models, such as ChatGPT, handle nonsensical sentences. The research found that even sophisticated models can mistake nonsensical sentences for meaningful ones, revealing limitations in their language processing capabilities. This highlights the need for further development to ensure AI systems can accurately discern and handle nonsensical input, which is crucial for their reliability in real-world applications.
* <https://pubmed.ncbi.nlm.nih.gov/39059084/> - A study published in the International Journal of Medical Informatics evaluated the response quality and readability of chatbots in the fields of cardiovascular health, oncology, and psoriasis. The research found significant variations in chatbot responses, with some deviating from established guidelines and adding extraneous information. This variability underscores the importance of carefully shaping chatbot responses, particularly in healthcare settings, to ensure accuracy and clarity for users seeking medical information.
* <https://medinform.jmir.org/2024/1/e54345> - A study published in JMIR Medical Informatics developed a Reference Hallucination Score (RHS) to evaluate the authenticity of AI chatbots' outputs. The research found significant variations in hallucination levels across different AI chatbots, with some exhibiting critical levels of hallucination. This underscores the necessity for robust evaluation tools to improve the reliability of AI chatbots and highlights the importance of verifying their outputs, especially in medical research contexts.