INVESTIGATING THE IMPACT OF GENERATIVE-AI ON THE SPREAD OF HEALTHCARE MISINFORMATION

Halley Smith[1], Caden Godwin[1], Amro Khasawneh[1]

[1]Department of Industrial Engineering and Industrial Management, School of Engineering,
Mercer University, Macon GA

Corresponding author's Email: khasawneh_an@mercer.edu

Author Note: Halley Smith is a graduate researcher pursuing a degree in Technical Management at Mercer University School of Engineering. Caden Godwin is also undergraduate research in Industrial Engineering at Mercer University as well. Dr. Amro Khasawneh is an Assistant Professor in the Department of Industrial Engineering and Industrial Management at Mercer University. His research focuses on patient safety, online decision-making, and human–machine interaction. Correspondence concerning this article should be addressed to Dr. Khasawneh at khasawneh_an@mercer.edu.

Abstract: As generative AI platforms become increasingly integrated into health information delivery, concerns about the accuracy and trustworthiness of AI-generated content continue to grow. This study explores how lay users evaluate and respond to dental health advice presented as either AI-generated or expert-generated. Participants reviewed responses to common dental questions and rated them based on trust, perceived accuracy, likelihood of following the advice, and confidence in the decision. Open-ended responses were analyzed using thematic analysis, revealing five key areas for improving AI interfaces: source citation, professional tone, credibility indicators, user personalization, and skepticism toward traditional providers. While many participants were open to AI-based health information, trust was conditional and influenced by transparency and presentation. These findings highlight actionable design features that can improve user confidence in AI health tools and offer critical insights into public perceptions of AI in healthcare.

Keywords: Healthcare misinformation, Patient trust, Artificial intelligence in dentistry, Human–AI interaction, Medical

decision-making, User perceptions of AI

1. Introduction

A recent poll showed that about %55 of U.S. adults rely on the Internet and social media to keep up to date on the news and for health information advice (Lopes et al., 2023). However, only a few of them would fully trust the information they receive from these platforms. At least %40 of Americans hear healthcare misinformation regularly and most of them are unsure whether the information is true or false (Lopes et al., 2023). A significant majority of U.S. adults believe that the spread of false and inaccurate information, both in general (86%) and specifically related to health issues (74%), is a major problem (Lopes et al., 2023). With the recent advances in artificial intelligence (AI), we expect this problem to become even more prominent.

While there is general excitement about the new opportunities brought by recent advancements AI, there is also a rise in patient exposure to medical misinformation (Zhou et al., 2023). We are now in an era of synthetic media, where text, images, audio, and video can be created or altered by generative AI models based on their training data. The commercial use of automated content from generative AI models, including large language models (LLMs) like ChatGPT and GPT-3, as well as image generation models, is rapidly expanding. While a few studies showed that AI-generated information can be accurate, due to the nature of how these models are built and depends on the training data (Goodman et al., 2023; Pan et al., 2023), these models can be unreliable. Thus, making errors in both facts and reasoning, which can be disseminated on an unprecedented scale (Amaro et al., 2023). The general public can easily obtain incorrect information from generative AI on any topic, including medicine (Moulaei et al., 2024). The spread of misinformation created by generative AI can be worsened by unsuspecting acceptance of the content's accuracy. There are serious potential negative consequences of medical misinformation for both individual care and public health (Moulaei et al., 2024). Therefore, it is important for us to understand the nature of the rapid spread of misinformation online and people behavior towards this information.

Traditional AI focuses on developing predictive models to carry out specific tasks, such as estimating numbers, classifying data, or choosing between a set of options. In contrast, generative AI aims to create original content. Instead of providing one correct answer based on decision boundaries, generative AI models generate text, audio, and visual outputs that can often be mistaken for human-produced content. Generative AI models rely on large neural networks trained with vast

amounts of raw data. The raw data that drives these generative AIs is human-produced and often coming from unverified/unknown sources to the user (Monteith et al., 2024). The raw training data is usually very large and constantly updated, making it very difficult to verify its accuracy (Ahmad et al., 2021). For instance, GPT-3 was trained on approximately 45 terabytes of text data, equivalent to about a million feet of bookshelf space (Monteith et al., 2024). Therefore, increasing the chances of AI producing misinformation.

Although the outputs from AI models can appear remarkably human-like, they often make reasoning and factual errors (Ahmad et al., 2021; Monteith et al., 2024). Researchers in the field refer to instances where the output is nonsensical, inaccurate, unfaithful to the underlying content, misleading, or incorrect as "hallucinations" (Salvagno et al., 2023). Generative AI models can produce various errors, including factual inaccuracies, inappropriate or dangerous advice, nonsense, fabricated sources, and arithmetic mistakes (Salvagno et al., 2023). Additionally, the responses can be outdated, reflecting the year the model was trained, and may vary with different iterations of the same question. An example of inappropriate advice is a chatbot suggesting calorie restriction and dieting to a user who disclosed having an eating disorder. Therefore, human judgment plays a crucial role in assessing the accuracy of generative AI outputs. It's common for people to underestimate the frequency of errors in generative AI models (Monteith et al., 2024; Roychowdhury, 2024; Salvagno et al., 2023). Unless one possesses expertise in the field, it's essential to thoroughly verify answers, even if they sound highly persuasive.

To date, there is limited research on how users interact with and respond to AI generated medical information. For example, Nadarzynski et al. (2019) studied patients' willingness to interact with an AI-led health chatbot through semi-structured interviews and found that most people are willing to accept information for these bots (Nadarzynski et al., 2019). On the other hand, Shinners et al. (2020) found that providers are usually hesitant to rely of AI-generated information unless they verify the source of such information (Shinners et al., 2020). Pan et al. (2023) and Goodman et al (2023) both studied providers reaction to AI-generated medical 2 contents. Both studies found that providers are generally satisfied with the AI-generated answers, but most providers thought the content was hard to understand by patients and often not readily actionable (Goodman et al., 2023; Pan et al., 2023). There is no research that qualitatively investigated patients' reactions to AI-generated medical information, their trust level, and ability to identify accurate verses inaccurate information.

2. Method

We used dentistry as it represents a familiar healthcare domain used by the majority of the population in the US (Ponathil et al., 2020). We asked 10-15 most commonly asked questions regarding dentistry to multiple generative-AI platforms and recorded the answers. We asked the generative-AI to provide wrong answers as well. The answers were then rated by healthcare experts regarding their accuracy. We used these answers to create a Type Form study that has both AI-generated and healthcare expert answers. Participants recruited through Type Form were asked to assume that they were looking for answers for the 9 questions and received these answers from two sources, (1) a dentist and (2) a generative-AI. Participants rated each response based on accuracy, trust in the information, the likelihood of following the advice, and confidence in their decision. Figure 1. Shows an example of a question and an AI response as provided in the study.
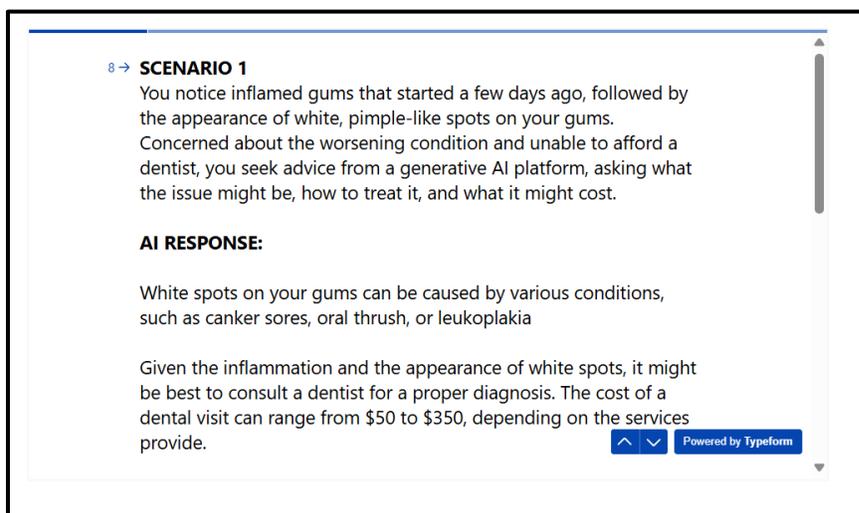

Figure 1. Scenario example with an AI response.

2.1 Participants

Table 1 shows the demographics of the participants. There are a wide range of people that participated. This is shown above in the table. There are 7 different categories that portray the number and percentage of each different type of participant. Gender shows that there were 30% male and 70% female participants. Race shows that the most participants were White/Caucasian at 62.5%. 20-29 was the biggest age range. The top 2 educations were high school diploma and bachelor's degree. The majority of the participants either had monthly or weekly experience with AI. 60% of those people did not have health care. A lot of people with AI experience and health care had a lower frequency of use with less than a month of experience.

Table 1. Participants Demographics.

	Number	%
Gender		
Male	12	30
Female	28	70
Race		
White/Caucasian	25	62.5
African American	6	15
Asian	4	10
American Indian or Alaska Native	1	2.5
Age		
20-29	20	50
30-39	10	25
40-59	7	17.5
60-79	3	7.5
Education		
High School Diploma	10	25
No Degree	2	5
Some College	8	20
Associates Degree	4	10
Bachelor's Degree	10	25
Masters Degree	4	10
PhD/Doctorate or equivalent	2	5
AI Experience		
Never	9	22.5
Less than once a month	4	10
Monthly	10	25
Weekly	10	25
Daily	7	17.5

2.2 Experimental Design

The study used a 2*2 within subject's experimental design, with the conditions being randomly ordered for each participant. The two independent variables for the study are following:

1. Accuracy of the answer: the answers have two levels, accurate and inaccurate. We asked the generative-AI to produce an accurate and inaccurate response to each question (4 answers total). However, we relied on the healthcare experts' categorization of the responses and the AI's categorization. One healthcare expert read each answer and picked out the two best answers. The healthcare expert also briefly described how they would best answer the question.

2. Source of the information: source of information variable also has two levels, dentist and AI-generated response. For each question the participants received an accurate answer that will be presented as coming from a dentist and another as coming from an AI.

After the participants answered the four dependent variable questions, they were asked to answer some extended response questions. The questions are:

1. What factors influence your decision to trust the response from an AI system over a dentist's opinion? Or vis versa?
2. Are there any communication styles or features that AI could adopt to seem more credible to you?
3. What would make you feel confident about choosing an AI response over a dentist's opinion in the future?
4. If you could give one piece of advice to the developers of the AI system, what would it be to ensure trustworthiness in its responses?

2.3 Procedure

At the start of the study, each participant will read an informed consent form to indicate their agreement to participate in the study. The participants were then provided with a demographic survey to fill up. The demographic survey includes questions about their age, education level, healthcare experience, and previous use of generative-AI. Upon successful completion of the demographic survey, participants were asked to assume that they were looking for dentistry advice, and they were given a scenario and the answer to the scenario. After reading each response, the participants answered the four dependent variables regarding their opinion of the response in the "experiment design" section above.

2.4 Data Analysis

We used thematic analysis to examine participants' open-ended responses and identify common themes related to their trust in AI-generated dental information and suggestions for improving AI interfaces. This method was selected for its flexibility and suitability for exploring user perceptions (Braun & and Clarke, 2006).

Two researchers independently reviewed the responses and generated initial codes using an inductive approach, allowing themes to emerge directly from the data rather than being imposed beforehand. After initial coding, the researchers met to compare their codebooks and refine the codes into broader thematic categories. Discrepancies were discussed and resolved through consensus to ensure consistency.

From this analysis, we identified several AI interface improvement ideas, including the need for source transparency (e.g., citations), professional tone, clear indicators of credibility (e.g., badges or user ratings), and greater customization. Frequency counts were recorded to indicate how often each theme appeared across responses, providing insight into the most common user priorities.

3. Results

Open-ended responses were analyzed by two researchers and grouped into thematic categories that represent participants' preferences for improving AI-generated health information interfaces. Five dominant themes emerged:

1. Source Citation and Verifiability

The most common suggestion (52.5%) was that AI-generated answers should include links to credible sources, citations, or references to back up their claims. This would help participants verify the information and increase trust. For example, one participant said:

"I would only choose an AI response over a dentist's opinion if the AI provided well-cited, evidence-based information from reputable sources like dental associations or medical journals. Even then, I'd still trust a dentist more for personalized advice based on their hands-on experience and assessment."

2. Human Connection and In-Person Expertise

About 15% of participants preferred traditional dental advice due to the perceived value of face-to-face interaction. They emphasized the importance of seeing a real dentist who could conduct a physical exam and establish a personal connection. For instance, a participant said:

"An AI system has all the information Google had which is great, provides loads of different studies. Although when it comes to a dentist, they can physically see and sometimes, good can't actually understand what you actually mean"

3. Professional Language and Tone

Another 15% of participants wanted AI responses to sound more professional, use correct medical terminology, and adopt a tone that mirrors expert communication. Responses that sounded too casual or oversimplified were perceived as less trustworthy. For example, one participant mentioned:

"Make the responses more professional, just the facts and scenarios no slang etc."

Another participant mentioned:

"AI could adopt a formal and professional tone, provide well-structured explanations, and cite credible sources when needed. Also, being transparent about limitations and avoiding vague or overly confident claims would make responses more trustworthy."

4. <u>Indicators of Trustworthiness</u>

Several participants (15%) indicated that they would trust AI more if it displayed some form of validation such as "fact-checked" labels, user ratings, or peer review acknowledgments. Visual indicators like "verified answer" badges were suggested.

"The development teams for the AI software should provide a badge of certification for the AI trustworthiness."

5. <u>Skepticism Toward Human Providers</u>

Interestingly, 7.5% of participants reported trusting AI more than dentists due to past negative experiences. Some felt that dental providers had tried to upsell procedures or provided inconsistent advice, making AI seem more impartial.

"Personal bias; I've had experiences with dentists who have given me misleading information in the past, so at times I was more inclined to trust the AI response."

4. Discussion

This study adds to the growing body of research on public trust in AI-generated health information by exploring how users evaluate and respond to AI- versus expert-generated dental advice. Our findings suggest that while many participants are open to using AI for health-related inquiries, trust is not automatically granted. It is influenced by several key factors including transparency, tone, and past experiences.

One of the most consistent suggestions from participants was the need for source transparency, such as the inclusion of references or links to reputable organizations. This aligns with prior research showing that perceived credibility of online health information is significantly increased when content includes citations or institutional affiliations (Lucassen & Schraagen, 2011). Without such markers, users may remain skeptical, particularly in domains like healthcare where the stakes are high.

Participants also emphasized the importance of professional tone and language in building trust. Casual or overly simplistic AI responses were perceived as less credible—echoing findings by Yoseph et al. (2024), who reported that healthcare professionals often rated AI-generated content as accurate but difficult for patients to interpret or act upon. This suggests that tone and delivery are just as important as factual accuracy when it comes to user acceptance.

Interestingly, a subset of participants expressed greater trust in AI than in human providers, citing previous negative experiences with dentists, such as perceived dishonesty or unnecessary procedures. This reflects a broader societal shift where distrust in traditional healthcare systems may drive some individuals toward automated alternatives (Kim, 2016; Nadarzynski et al., 2019). However, this trust is conditional. Participants still wanted safeguards, such as ratings or visual indicators of accuracy (e.g., thumbs up/down, expert-reviewed badges).

The study also highlights a desire for user agency and customization, which is consistent with research in human–computer interaction showing that users are more likely to trust systems they can control or personalize (Luger & Sellen, 2016). Participants wanted to see AI interfaces that adapt to their preferences, offer varying levels of detail, and clearly indicate uncertainty or limitations in the information provided.

These insights have important implications for the design of AI-driven health communication tools. Developers should consider embedding credibility cues, providing options for explanation depth, and using language tailored to the target user's literacy level. Simple interface tweaks, such as including "last updated" timestamps, verified checkmarks, or source footnotes, could significantly improve user trust and engagement.

Finally, this study addresses a gap in the literature, as most existing research focuses on clinicians' views of AI (Dong et al., 2025; Shinners et al., 2020), while few studies investigate how laypeople interact with AI-generated medical advice. Our findings show that the general public uses a diverse set of heuristics, beyond factual accuracy, to evaluate trustworthiness, including tone, design, and perceived transparency.

5. References

Ahmad, Z., Rahim, S., Zubair, M., & Abdul-Ghafar, J. (2021). Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: Present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review. *Diagnostic Pathology*, *16*(1), 24. https://doi.org/10.1186/s13000-021-01085-4

Amaro, I., Della Greca, A., Francese, R., Tortora, G., & Tucci, C. (2023). AI Unreliable Answers: A Case Study on ChatGPT. In H. Degen & S. Ntoa (Eds.), *Artificial Intelligence in HCI* (pp. 23–40). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-35894-4_2

Braun, V., & and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Dong, C., Aw, D. C. W., Lee, D. W. C., Low, S. C., & Yan, C. C. (2025). Generative AI in Healthcare: Insights from Health Professions Educators and Students. *International Medical Education*, *4*(2), Article 2. https://doi.org/10.3390/ime4020011

Goodman, R. S., Patrinely, J. R., Stone, C. A., Jr, Zimmerman, E., Donald, R. R., Chang, S. S., Berkowitz, S. T., Finn, A. P., Jahangir, E., Scoville, E. A., Reese, T. S., Friedman, D. L., Bastarache, J. A., van der Heijden, Y. F., Wright, J. J., Ye, F., Carter, N., Alexander, M. R., Choe, J. H., … Johnson, D. B. (2023). Accuracy and Reliability of Chatbot Responses to Physician Questions. *JAMA Network Open*, *6*(10), e2336483. https://doi.org/10.1001/jamanetworkopen.2023.36483

Kim, Y. (2016). Trust in health information websites: A systematic literature review on the antecedents of trust. *Health Informatics Journal*, *22*(2), 355–369. https://doi.org/10.1177/1460458214559432

Lopes, L., Kearney, A., Washington, I., Valdes, I., Yilma, H., & Published, L. H. (2023, August 22). KFF Health Misinformation Tracking Poll Pilot. *KFF*. https://www.kff.org/coronavirus-covid-19/poll-finding/kff-health-misinformation-tracking-poll-pilot/

Lucassen, T., & Schraagen, J. M. (2011). Factual accuracy and trust in information: The role of expertise. *Journal of the American Society for Information Science and Technology*, *62*(7), 1232–1242. https://doi.org/10.1002/asi.21545

Luger, E., & Sellen, A. (2016). "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297. https://doi.org/10.1145/2858036.2858288

Monteith, S., Glenn, T., Geddes, J. R., Whybrow, P. C., Achtyes, E., & Bauer, M. (2024). Artificial intelligence and increasing misinformation. *The British Journal of Psychiatry*, *224*(2), 33–35. https://doi.org/10.1192/bjp.2023.136

Moulaei, K., Yadegari, A., Baharestani, M., Farzanbakhsh, S., Sabet, B., & Reza Afrash, M. (2024). Generative artificial intelligence in healthcare: A scoping review on benefits, challenges and applications. *International Journal of Medical Informatics*, *188*, 105474. https://doi.org/10.1016/j.ijmedinf.2024.105474

Nadarzynski, T., Miles, O., Cowie, A., & Ridge, D. (2019). Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *DIGITAL HEALTH*, *5*, 2055207619871808. https://doi.org/10.1177/2055207619871808

Pan, A., Musheyev, D., Bockelman, D., Loeb, S., & Kabarriti, A. E. (2023). Assessment of Artificial Intelligence Chatbot Responses to Top Searched Queries About Cancer. *JAMA Oncology*, *9*(10), 1437–1440. https://doi.org/10.1001/jamaoncol.2023.2947

Ponathil, A. P., Khasawneh, A., Byrne, K., & Madathil, K. C. (2020). Factors affecting the choice of a dental care provider by older adults based on online consumer reviews. *IISE Transactions on Healthcare Systems Engineering*, 1–19. https://doi.org/10.1080/24725579.2020.1854394

Roychowdhury, S. (2024). Journey of Hallucination-minimized Generative AI Solutions for Financial Decision Makers. *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, 1180–1181. https://doi.org/10.1145/3616855.3635737

Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Artificial intelligence hallucinations. *Critical Care*, *27*(1), 180. https://doi.org/10.1186/s13054-023-04473-y

Shinners, L., Aggar, C., Grace, S., & Smith, S. (2020). Exploring healthcare professionals' understanding and experiences of artificial intelligence technology use in the delivery of healthcare: An integrative review. *Health Informatics Journal*, *26*(2), 1225–1236. https://doi.org/10.1177/1460458219874641

Yoseph, E. T., Gonzalez-Suarez, A. D., Lang, S., Desai, A., Hu, S. S., & Zygourakis, C. C. (2024). Patient perspectives on AI: A pilot study comparing large language model and physician-generated responses to routine cervical spine surgery questions. *Artificial Intelligence Surgery*, *4*(3), 267–277. https://doi.org/10.20517/ais.2024.38

Zhou, J., Zhang, Y., Luo, Q., Parker, A. G., & De Choudhury, M. (2023). Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–20. https://doi.org/10.1145/3544548.3581318