

Article

Leveraging Large Language Models to Address Common Vaccination Myths and Misconceptions

Florian Reis *, Lea J. Bayer, Claudius Malerczyk, Christian Lenz and Christof von Eiff

Medical Affairs, Pfizer Pharma GmbH, Berlin 10117, Germany

* Correspondence: florian.reis@pfizer.com

Abstract

Background/Objectives: Large language models (LLMs) are increasingly used by the public to seek health information, yet their accuracy in addressing common vaccine myths remains unclear. Sycophantic LLM behavior, where models align with rather than correct user-stated beliefs, poses specific risks in health contexts. **Methods:** We conducted an exploratory multi-vendor evaluation of three LLMs (GPT-5, Gemini 2.5 Flash, Claude Sonnet 4) using officially curated vaccination myths from Germany's public health institution and two realistic user framings (curious skeptic, convinced believer). All model responses were independently evaluated by two blinded medical experts for misconception addressal (binary criterion applied to the response text), scientific accuracy, and communication clarity (5-point Likert scales). Additionally, blinded marketing experts ranked models for lay communication clarity. Flesch Reading Ease scores were computed for all outputs. **Results:** Across all myths, framings, and models (66 response items), both medical raters judged that all responses refuted the targeted misconception; no response affirmed or ignored a myth, including under the adversarial convinced believer framing. Scientific accuracy and clarity ratings were high and tightly clustered (median 4.0–4.5), with no combined score below 3 and substantial inter-rater agreement. Marketing experts independently ranked Gemini 2.5 Flash and GPT-5 highest for lay clarity. Readability analysis revealed generally low accessibility, particularly for the convinced believer framing and for Claude Sonnet 4 outputs. **Conclusions:** Our findings suggest that general-purpose LLMs can produce scientifically accurate, on-topic rebuttals to widely documented vaccine myths under realistic default conditions, although linguistic complexity and framing-sensitive style may limit accessibility. Whether such outputs change beliefs or behavior in hesitant individuals was not tested. With readability optimization, these outputs could serve as building blocks for myth-debunking tools, given prospective evaluation with behavioral endpoints.

Academic Editor: Pedro Plans-Rubió

Received: 2 June 2026

Revised: 30 June 2026

Accepted: 3 July 2026

Published: 3 July 2026

Keywords: vaccine hesitancy; large language models; vaccine myths; health literacy; health communication; myth debunking

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Vaccination is among the most effective public health interventions, increasing life expectancy and reducing the burden of infectious diseases worldwide [1,2]. Despite proven benefits, vaccine hesitancy remains a persistent challenge, fueled in part by the widespread circulation of misinformation [3,4]. Myths and misconceptions about vaccine

safety and efficacy continue to undermine public trust, leading to lower vaccination rates and increased vulnerability to preventable diseases [5]. Consequently, the World Health Organization has identified vaccine hesitancy as a leading threat to global health [6], underscoring the urgency of effective countermeasures.

In recent years, the rise of artificial intelligence (AI) and, in particular, large language models (LLMs) has transformed how people can access health information [7]. As these models are widely available to the public, their ability to instantly provide fluent, personalized, and seemingly qualified responses offers a convenient alternative to traditional information sources such as publications and public health agency reports. However, an unregulated and unsupervised deployment of LLMs raises concerns about the accuracy and safety of the content they generate [8]. Inaccurate or misleading responses, particularly on health-related topics, may inadvertently amplify misinformation rather than correct it, thus posing a threat to public health [9]. The dual role of LLMs as both potential debunkers and amplifiers of misinformation further underscores the importance of their reliability in generating health advice. While these models have demonstrated impressive performance in various domains of healthcare [7,10], their ability to address complex and sensitive topics such as vaccination myths and misconceptions remains underexplored. To date, only limited empirical evidence exists on how LLMs perform in this context [11–13]. The few available studies frequently lack comparisons across different vendors, do not draw from professionally curated sources or disregard the influence of prompt framing on model outputs.

2. Materials and Methods

2.1. Study Design

Our study employed a cross-model comparative design to evaluate the performance of LLMs in addressing common vaccination myths and misconceptions. Specifically, we examined the effect of prompt framing and compared the performance of different models across multiple rater consortia (Figure 1).

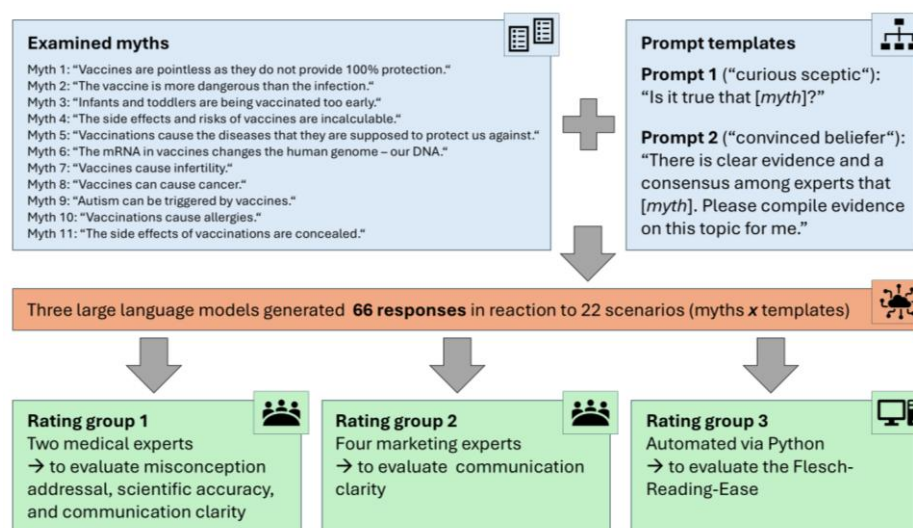


Figure 1. Graphical illustration of the experimental setup and the examined vaccination myths and prompts.

The primary outcomes were misconception addressal, scientific accuracy, and communication clarity of LLM-generated responses. Eleven commonly circulated myths about vaccinations were selected based on a compendium compiled by the Robert Koch Institute, Germany's public health institution governing disease control and prevention

[14]. These myths cover key domains such as vaccine efficacy, safety, risk–benefit balance, and transparency (see Supplementary Table S1).

2.2. Examined Models and Prompts

For this study, three popular state-of-the-art LLMs were evaluated: OpenAI GPT-5, Google Gemini 2.5 Flash, and Anthropic Claude Sonnet 4. We selected these systems as prominent, broadly accessible, general-purpose models that were highly visible in the public domain and commonly represented among leading models in benchmark comparisons and leaderboards at the time of data collection. The study was designed to compare a focused set of widely used systems rather than to provide an exhaustive benchmark of all available models. Two distinct prompt templates were applied to each myth to simulate different states of user belief: “curious skeptic” (prompt: “Is it true that [myth]?”) and “convinced believer” (prompt: “There is clear evidence and a consensus among experts that [myth]. Please compile evidence on this topic for me.”). All prompts were entered via the models’ standard web interfaces in July 2025, using the options offered free of charge by each provider (within standard usage limits) with default settings, to simulate layperson use. Chat histories were cleared before and during each iteration to prevent personalization and memory effects.

2.3. Rating Approach

To ensure a comprehensive assessment, we formed three independent rater cohorts. The first cohort was a medical group to evaluate the primary outcomes; the second cohort comprised marketing experts focusing on communication clarity to laypeople; the third rating approach comprised an automated calculation of the Flesch Reading Ease. Two medical raters with MD or PhD qualifications and over 20 years of combined vaccine-related medical expertise independently evaluated all LLM-generated responses. Both raters were blinded to the model identities.

We matched each measure to the nature of the underlying construct. Misconception addressal is a categorical, health-related and thus AI safety-relevant property of a response, and was therefore captured as a single binary criterion presented identically to both medical raters: whether “the misinformation is explicitly addressed and refuted” by the response (rated yes/no). This criterion concerns the content of the generated response as judged by experts and does not measure any change in user beliefs. Scientific accuracy and communication clarity are graded constructs and were therefore rated on ordinal 5-point Likert scales. Both raters applied all criteria independently; their agreement is reported in the Results Section. In total, this approach resulted in 396 observation items ($11 \text{ myths} \times 2 \text{ prompts} \times 3 \text{ models} \times 3 \text{ measurements} \times 2 \text{ raters}$).

To further analyze communication clarity from a layperson’s perspective, four expert raters with pharmaceutical marketing expertise independently assessed all model outputs. Blinded to the model identities and the type of prompt, the raters were asked to directly compare and rank model performance regarding communication clarity towards laypeople. Because the medical Likert ratings were uniformly high and tightly clustered (a ceiling effect that limited between-model discrimination), we deliberately adopted a forced-choice ranking for this layperson-clarity assessment (see Supplementary Table S2). Finally, to also provide objective metrics, we calculated Flesch Reading Ease scores using the textstat library in Python to quantify the readability of the respective model outputs. Provided source links and display items were excluded from this text-based calculation. Flesch Reading Ease scores between 31 and 50 are considered difficult to read (college level), while scores between 0 and 30 are considered very difficult (college graduate level) [15].

2.4. Statistical Analysis

All analyses were conducted in Python (v3.12). Due to the exploratory nature of the study design, we consider the results to be descriptive and do not conduct hypothesis testing. For inter-rater reliability between the two medical raters (ordinal data), we report Gwet's AC2 with quadratic weights.

3. Results

Our findings in the medical rater group revealed that all three LLMs correctly addressed all eleven myths regardless of the prompt used ("misconception addressal", binary rating). Averaged ratings of both medical raters for "scientific accuracy" and "communication clarity" (5-point Likert scale) were high and tightly clustered for all models and across both prompts. For scientific accuracy, medians were 4.5 for both prompts, with narrow interquartile ranges (IQRs: prompt 1: 4.0–4.5; prompt 2: 4.0–4.5). No combined score fell below 3. For prompt 1, 97% of items were ≥ 4 ($70\% \geq 4.5$). For prompt 2, 91% were ≥ 4 ($70\% \geq 4.5$). For communication clarity, medians were 4.5 for prompt 1 and 4.0 for prompt 2 (IQRs: prompt 1, 4.5–4.5; prompt 2, 3.5–4.5). No combined score fell below 3. For prompt 1, 97% were ≥ 4 ($76\% \geq 4.5$). For prompt 2, 70% were ≥ 4 ($36\% \geq 4.5$) (Figure 2).

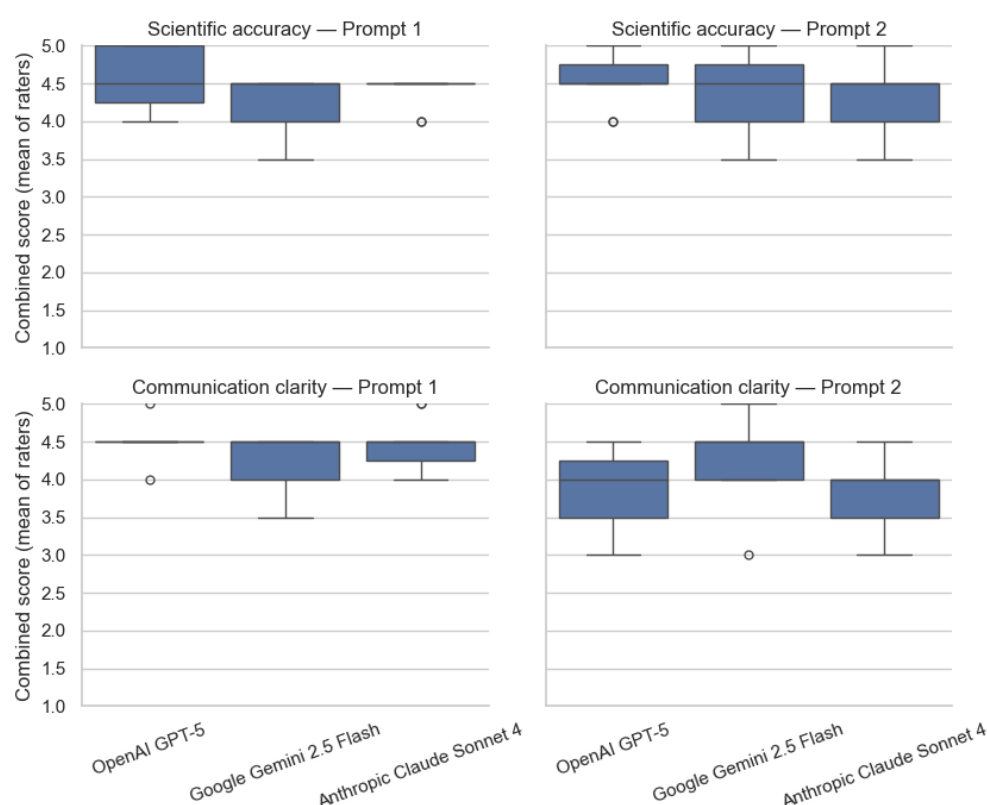


Figure 2. Combined medical ratings by LLM and prompt for scientific accuracy and communication clarity (mean of two raters). Prompt 1: "curious skeptic"; prompt 2: "convinced believer".

We observed 100% agreement between both medical raters for "misconception addressal". For "scientific accuracy" and "communication clarity", we found strong practical alignment between both raters, as most ratings were within one point (scientific accuracy: 83%; communication clarity: 76%). As these ordinal data were skewed and highly concentrated, traditional coefficients (e.g., kappa/ICC) were low and not suitable. Therefore, we calculated chance-corrected agreement using Gwet's AC2 with quadratic

weights, which was substantial for both domains: scientific accuracy AC2 = 0.82 (95% CI: 0.76–0.89) and communication clarity AC2 = 0.72 (95% CI: 0.62–0.83).

Regarding our additional analysis of communication clarity towards laypeople, which was independently rated by marketing experts, we asked for comparative model rankings from 1 to 3 (lower is better). Because the medical ratings were uniformly high and tightly clustered, we used this forced-choice ranking approach to enable a more nuanced, discriminating evaluation. On average, the models were ranked as follows: Google Gemini 2.5 Flash came first with a mean rank of about 1.83, OpenAI GPT-5 came second with a mean rank of about 1.93, and Anthropic Claude Sonnet 4 came third with a mean rank of about 2.24. Analyzing the prompts revealed where the discrepancies originate. For prompt 1, GPT-5 and Gemini performed similarly, with GPT-5 achieving a score of around 1.84 and Gemini achieving a score of around 1.95, while Claude fell behind with a score of around 2.20. For prompt 2, Gemini led with a score of around 1.70, while GPT-5 fell behind with a score of around 2.02, and Claude remained steady at around 2.27 (see Supplementary Materials for detailed analysis).

Additionally, we calculated the Flesch Reading Ease score based on word count, sentences, and syllables in the outputs of all three models. This revealed lower readability scores for all models when using prompt 2 (“convinced believer”). Overall, the readability score for GPT-5 was the highest with 35.9 for prompt 1 and 30.1 for prompt 2; for Gemini 2.5-Flash, it was 37.8 and 30.0; and for Claude Sonnet-4, it was 27.2 and 21.0, respectively.

4. Discussion

This explorative multi-vendor evaluation demonstrates that LLMs can reliably generate scientifically accurate, on-topic rebuttals to common vaccination myths when queried under realistic default settings. Across eleven myths and two user framings, all models consistently rejected misinformation in their responses. Medical expert ratings placed scientific accuracy and clarity in high ranges, with no substantial misinformation. We emphasize that these measures characterize the content of model outputs as judged by experts; they do not establish whether such outputs change the beliefs or decisions of users who encounter the myths. Our findings suggest that general-purpose LLMs can meet the quality threshold required for communication about widely documented vaccine myths, as outlined in the literature [11,13]. Our layered assessment, conducted by marketing experts and evaluated through a lay-communication lens, resulted in Gemini 2.5 Flash achieving the highest score, with GPT-5 close behind and Claude Sonnet 4 less favored. Differences were most pronounced for the “convinced believer” framing, where outputs became more complex and less readable across all models. Notably, the readability of the answers generated by the models was difficult. In Claude Sonnet 4, with Flesch Reading Ease scores below 30, the readability was very difficult and appropriate only for an academic audience. This indicates that while factual correctness is robust, accessibility overall remains challenging as well as model- and context-dependent, especially for users starting from strong misbelief.

Prior evaluations of LLMs as a counterweight to vaccine misinformation have reached broadly encouraging conclusions that our results build upon. Sohail et al. analyzed ChatGPT’s responses to vaccine-misconception queries and argued that such chatbots could help shape perceptions and reduce misconceptions, but their evaluation adopted a more illustrative, single-model perspective [11]. Koh et al. had two physicians judge ChatGPT’s answers to fifteen hesitancy questions against official guidance and likewise found consistently fact-based responses; notably, they did not report a formal scoring rubric but explicitly called for one [13]. Joshi et al. compared ChatGPT with an official source across English and Spanish, showing that information quality and linguistic consistency differed between the two languages [12]. Our study aligns with all three on the

central finding that contemporary general-purpose LLMs produce accurate responses to common vaccine myths, and echoes Joshi et al. in identifying accessibility, rather than factual correctness, as a critical constraint. However, our study differs from this literature in four ways. First, rather than evaluating one vendor, we benchmarked three models head-to-head. Second, instead of assuming a single neutral user, we contrasted different framings, testing explicitly for prompt sensitivity and the associated risk of sycophancy, a failure mode that single-prompt designs cannot surface. Third, our myths were drawn from an officially curated public-health compendium rather than author-generated questions. Fourth, we conducted the blinded, rubric-based evaluation that Koh et al. requested, which covered subjective measures, such as accuracy and clarity, as well as objective readability. Our findings thus extend prior work showing that LLMs often provide accurate vaccine-related information but vary in quality, readability, and alignment with public health guidance [16,17].

Varying prompt framing illustrates how subtle shifts in user input can change style and complexity without necessarily altering correctness, which is an essential consideration for real-world applications. Models may remain susceptible to misleading or overconfident statements or shift behavior over time, which underscores the need for explicit safeguards against “AI sycophancy” and a potential amplification of misconceptions. This aligns with recent publications that demonstrated the risks of sycophantic AI behavior [18] and how LLMs can produce less accurate output when trained to produce warmer responses [19]. Furthermore, it emphasizes the importance of considering not only the technical capabilities of innovative tools but also how users interact with them [20]. Even the most capable model can only process the information it is given, a particularly critical issue in high-stakes environments like healthcare.

These results have implications for digital public health, given that misinformation and organized anti-vaccine networks influence undecided audiences [3–5,21]. Integrated into official websites or patient portals, LLMs could in principle act as scalable myth-debunking assistants [13,22]. Evidence from conversational interventions against misinformation supports the promise of personalized, dialogic formats to improve vaccine-related outcomes and reduce conspiracy beliefs [23,24]. Importantly, those effects were established in dedicated intervention studies. Our study suggests that current models can already provide high-quality building blocks for such tools, if embedded within supervised and auditable systems. However, LLMs can still exhibit overconfidence, unstable behavior, prompt sensitivity, and difficulties with epistemic distinctions, raising concerns about real-world deployment [8,17,25]. Transparent sourcing, continuous monitoring, and alignment with regulatory and ethical standards are therefore necessary [8,22].

Strengths of this work include the use of officially curated myths, a multi-vendor comparison, manipulation of user framing, blinded multi-expert evaluation, and comparisons with objective readability metrics. Several limitations restrict the scope of our conclusions. First, the myths were succinct, officially curated items. We did not test rhetorically sophisticated misinformation or multi-turn exchanges. Second, our outcomes are expert judgments of output quality, not measures of persuasion. We did not assess attitude change, belief revision, or vaccination behavior among hesitant individuals. Third, our results reflect the specific model versions and access tiers that were available in mid-2025, so they may not be applicable to other providers, tiers, or future model updates. Other limitations of this exploratory study include the limited number of myths and the non-deterministic nature of model outputs.

5. Conclusions

In conclusion, general-purpose LLMs can produce expert-rated, scientifically accurate and on-topic rebuttals to prevalent vaccine myths. However, their linguistic accessi-

bility is limited and framing-dependent. These findings position LLMs as potentially promising elements of public-health communication when paired with transparent sourcing and readability optimization within supervised and auditable systems. Whether these outputs actually reduce public misconceptions or influence vaccination decisions remains an open question that requires prospective evaluation in real-world settings with laypeople and behavioral endpoints.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/vaccines14070594/s1>, Table S1: Overview of the Robert Koch Institute's "fact sandwiches" used as input for this study; Table S2: Model performance regarding communication clarity as rated by marketing experts.

Author Contributions: Conceptualization, F.R., C.L. and C.v.E.; Methodology, F.R.; Software, F.R.; Validation, L.J.B. and C.M.; Formal Analysis, F.R.; Data Curation, F.R.; Writing—Original Draft Preparation, F.R.; Writing—Review and Editing, C.v.E., L.J.B., C.M. and C.L.; Visualization, F.R.; Supervision, C.v.E. and C.L.; Project Administration, F.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: This study does not involve human study participants or personally identifiable human data.

Data Availability Statement: All medical and marketing ratings necessary to replicate the results of this study are openly available on Zenodo (<https://doi.org/10.5281/zenodo.17617792>). The full dataset of generated model responses is available from the corresponding author upon reasonable request. No custom algorithms were developed; all computations relied on publicly available Python libraries.

Acknowledgments: We wish to thank Katharina Beuchert, Paul Bienwald, Stephanie Klein, and Leonie Stroh for providing additional ratings for the case vignettes.

Conflicts of Interest: All authors are current employees of Pfizer, and hold company shares and stock options. Every interpretation and all opinions expressed in this article are those of the authors and not necessarily those of Pfizer. The authors state that there are no other conflicts of interest.

References

1. Andre, F.E.; Booy, R.; Bock, H.L.; Clemens, J.; Datta, S.K.; John, T.J.; Lee, B.W.; Lolekha, S.; Peltola, H.; Ruff, T.A.; et al. Vaccination greatly reduces disease, disability, death and inequity worldwide. *Bull. World Health Organ.* **2008**, *86*, 140–146. <https://doi.org/10.2471/blt.07.040089>.
2. Orenstein, W.A.; Ahmed, R. Simply put: Vaccination saves lives. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 4031–4033. <https://doi.org/10.1073/pnas.1704507114>.
3. Wilson, S.L.; Wiysonge, C. Social media and vaccine hesitancy. *BMJ Glob. Health* **2020**, *5*, e004206. <https://doi.org/10.1136/bmjgh-2020-004206>.
4. Betsch, C.; Brewer, N.T.; Brocard, P.; Davies, P.; Gaissmaier, W.; Haase, N.; Leask, J.; Renkewitz, F.; Renner, B.; Reyna, V.F.; et al. Opportunities and challenges of Web 2.0 for vaccination decisions. *Vaccine* **2012**, *30*, 3727–3733. <https://doi.org/10.1016/j.vaccine.2012.02.025>.
5. Dubé, E.; Laberge, C.; Guay, M.; Bramadat, P.; Roy, R.; Bettinger, J.A. Vaccine hesitancy: An overview. *Hum. Vaccines Immunother.* **2013**, *9*, 1763–1773. <https://doi.org/10.4161/hv.24657>.
6. WHO. Ten Threats to Global Health in 2019. Available online: <https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019> (accessed on 1 July 2025).
7. Thirunavukarasu, A.J.; Ting, D.S.J.; Elangovan, K.; Gutierrez, L.; Tan, T.F.; Ting, D.S.W. Large language models in medicine. *Nat. Med.* **2023**, *29*, 1930–1940. <https://doi.org/10.1038/s41591-023-02448-8>.

8. Meskó, B.; Topol, E.J. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *npj Digit. Med.* **2023**, *6*, 120. <https://doi.org/10.1038/s41746-023-00873-0>.
9. Ruggeri, K.; Vanderslott, S.; Yamada, Y.; Argyris, Y.A.; Večkalov, B.; Boggio, P.S.; Fallah, M.P.; Stock, F.; Hertwig, R. Behavioural interventions to reduce vaccine hesitancy driven by misinformation on social media. *BMJ* **2024**, *384*, e076542. <https://doi.org/10.1136/bmj-2023-076542>.
10. Lee, P.; Bubeck, S.; Petro, J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. *N. Engl. J. Med.* **2023**, *388*, 1233–1239. <https://doi.org/10.1056/nejmsr2214184>.
11. Sohail, S.S.; Madsen, D.Ø.; Farhat, F.; Alam, M.A. ChatGPT and Vaccines: Can AI Chatbots Boost Awareness and Uptake? *Ann. Biomed. Eng.* **2023**, *52*, 446–450. <https://doi.org/10.1007/s10439-023-03305-y>.
12. Joshi, S.; Ha, E.; Amaya, A.; Mendoza, M.; Rivera, Y.; Singh, V.K. Ensuring Accuracy and Equity in Vaccination Information From ChatGPT and CDC: Mixed-Methods Cross-Language Evaluation. *JMIR Form. Res.* **2024**, *8*, e60939. <https://doi.org/10.2196/60939>.
13. Koh, M.C.Y.; Ngiam, J.N.; Salada, B.M.A.; Tambyah, P.A.; Archuleta, S.; Oon, J.E.L. Can ChatGPT Counter Vaccine Hesitancy? An Evaluation of ChatGPT's Responses to Simulated Queries from the General Public. *Healthcare* **2025**, *13*, 1269. <https://doi.org/10.3390/healthcare13111269>.
14. RKI. Vaccination Myths: Effectively Debunking Misinformation. 2024. Available online: <https://www.rki.de/EN/Topics/Infectious-diseases/Immunisation/Information-material/Vaccination-myths/effectively-debunking-misinformation-node.html> (accessed on 1 July 2025).
15. Hong, S.W.; Kang, J.-H.; Park, J.H.; Park, H.J.; Kim, E. Quality and readability of online information on hand osteoarthritis. *Heal. Inform. J.* **2023**, *29*, 14604582231169297. <https://doi.org/10.1177/14604582231169297>.
16. Pan, A.; Musheyev, D.; Bockelman, D.; Loeb, S.; Kabarriti, A.E. Assessment of Artificial Intelligence Chatbot Responses to Top Searched Queries About Cancer. *JAMA Oncol.* **2023**, *9*, 1437–1440. <https://doi.org/10.1001/jamaoncol.2023.2947>.
17. Rebitschek, F.G.; Carella, A.; Kohlrausch-Pazin, S.; Zitzmann, M.; Steckelberg, A.; Wilhelm, C. Evaluating evidence-based health information from generative AI using a cross-sectional study with laypeople seeking screening information. *npj Digit. Med.* **2025**, *8*, 343. <https://doi.org/10.1038/s41746-025-01752-6>.
18. Cheng, M.; Lee, C.; Khadpe, P.; Yu, S.; Han, D.; Jurafsky, D. Sycophantic AI decreases prosocial intentions and promotes dependence. *Science* **2026**, *391*, eaec8352. <https://doi.org/10.1126/science.aec8352>.
19. Ibrahim, L.; Hafner, F.S.; Rocher, L. Training language models to be warm can reduce accuracy and increase sycophancy. *Nature* **2026**, *652*, 1159–1165. <https://doi.org/10.1038/s41586-026-10410-0>.
20. Reis, M.; Reis, F.; Kim, Y.J.; Demir, A.; Lim, J.; Gröschel, M.I.; Boie, S.D.; Kunde, W. Reduced symptom reporting quality during human–chatbot versus human–physician interactions. *Nat. Health* **2026**. <https://doi.org/10.1038/s44360-026-00116-y>.
21. Johnson, N.F.; Velásquez, N.; Restrepo, N.J.; Leahy, R.; Gabriel, N.; El Oud, S.; Zheng, M.; Manrique, P.; Wuchty, S.; Lupu, Y. The online competition between pro- and anti-vaccination views. *Nature* **2020**, *582*, 230–233. <https://doi.org/10.1038/s41586-020-2281-1>.
22. Passanante, A.; Pertwee, E.; Lin, L.; Lee, K.Y.; Wu, J.T.; Larson, H.J. Conversational AI and Vaccine Communication: Systematic Review of the Evidence. *J. Med. Internet Res.* **2023**, *25*, e42758. <https://doi.org/10.2196/42758>.
23. Wang, Z.; Chen, S.; Poon, J.; Han, S.C.; Ye, D.; Yu, F.-Y.; Fang, Y.; Ni, Z.; Wong, M.C.S.; Mo, P.K.H. A Hybrid Chatbot to Promote Pneumococcal Vaccination Among Older Adults. *JAMA Netw. Open* **2025**, *8*, e2535813. <https://doi.org/10.1001/jamanetworkopen.2025.35813>.
24. Costello, T.H.; Pennycook, G.; Rand, D.G. Durably reducing conspiracy beliefs through dialogues with AI. *Science* **2024**, *385*, eadq1814. <https://doi.org/10.1126/science.adq1814>.
25. Suzgun, M.; Gur, T.; Bianchi, F.; Ho, D.E.; Icard, T.; Jurafsky, D.; Zou, J. Language models cannot reliably distinguish belief from knowledge and fact. *Nat. Mach. Intell.* **2025**, *7*, 1780–1790. <https://doi.org/10.1038/s42256-025-01113-8>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.