Can Large Language Models Reduce Vaccine Conspiracy Beliefs?

Evgeniya Hristova (ehristova@cogs.nbu.bg) & Maurice Grinberg
Research Center for Cognitive Science & Department of Cognitive Science and Psychology, New Bulgarian University

Abstract

Vaccine conspiracy beliefs contribute to vaccine hesitancy and resistance to public health interventions. Recent work suggests that large language models (LLMs) may reduce conspiracy beliefs through personalized, evidence-based dialogue. The present study extends this work by examining whether LLM-mediated conversations reduce vaccine conspiracy beliefs and whether effectiveness depends on conversational style in a country characterized by negative attitudes towards vaccination. Participants engaged in an interactive conversation with an LLM following open-ended elicitation of their vaccine conspiracy beliefs. The LLM used rational, emotional, or free-argumentation, or a neutral control conversation. Belief in vaccine conspiracies was assessed before and after the interaction using summaries of the participant's personally endorsed vaccine conspiracy beliefs. Results showed a reduction in vaccine conspiracy beliefs only when rational argumentation is used. However, no effect was observed for vaccine-related behavioral intentions. These findings confirm the conclusions of previous studies that belief change via AI requires rational engagement.

Keywords: vaccine conspiracy beliefs; vaccine hesitancy; large language models; conversational AI; misinformation; belief revision

Introduction

Vaccine Conspiracy Beliefs

Vaccine conspiracy theories involve the belief that vaccines are intentionally harmful or that information about their safety and effectiveness is deliberately concealed by powerful actors such as governments, pharmaceutical companies, or medical authorities. A growing body of research shows that endorsement of these beliefs is associated with greater institutional mistrust, heightened perceptions of vaccine risk, rejection of scientific consensus, lower vaccination intentions, increased safety concerns, and resistance to corrective information (Jolley & Douglas, 2014; Taubert et al., 2024).

Vaccine-related conspiracy narratives are widely present online and are endorsed by a considerable part of the population, with prevalence varying considerably across countries, vaccine types, and historical contexts (Taubert et al., 2024). These narratives are particularly prominent on social media platforms and became especially widespread during the COVID-19 pandemic. Across studies, belief endorsement ranges from very low levels (about 2%) to extremely high levels (over 70%), depending on the specific narrative and population examined (Taubert et al., 2024).

Vaccine hesitancy – defined as the delay in acceptance or refusal of vaccines despite availability – is strongly and consistently linked to conspiracy beliefs. Evidence from numerous correlational studies indicates that individuals who endorse vaccine conspiracy narratives report lower vaccination intentions and lower vaccine uptake (Steinert et al., 2022; Taubert et al., 2024).

Steinert et al. (2022) examined COVID-19 vaccine hesitancy across eight European countries, focusing on the role of misinformation and conspiracy beliefs as contextual determinants of both hesitancy and responsiveness to public health messaging. Using survey data from more than 10,000 adults collected in 2021, the authors documented substantial cross-national variation in hesitancy, ranging from approximately 6% in Spain to over 60% in Bulgaria.

Vaccination rates in Bulgaria are particularly low compared with other EU countries for vaccines that are not included in the mandatory childhood immunisation calendar. Only 30.4% of the Bulgarian population had completed the initial COVID-19 vaccination protocol, compared with 73.5% in the European Union (Our World in Data, 2026). OECD data show that Bulgaria has one of the lowest HPV vaccination rates in Europe. In 2023, an average of 64% of girls in the European Union had received all recommended HPV vaccine doses by age 15, whereas coverage in Bulgaria was only 7%, the lowest rate among the countries included in the comparison (OECD, 2024).

UNICEF Bulgaria and the Bulgarian Ministry of Health (2023) conducted a nationally representative study of parental attitudes toward childhood immunization in Bulgaria. The study indicates that vaccine hesitancy is closely linked to misinformation, distrust, and conspiracy-related beliefs.

Large Language Models in the Reduction of Conspiracy Beliefs

The substantial public health consequences of vaccine hesitancy raise the question of whether effective and scalable interventions can reduce conspiracy beliefs or their downstream effects. Taubert et al. (2024) identified a small but growing body of experimental research testing strategies to reduce conspiracy beliefs or increase vaccination intentions. However, effect sizes were consistently small, and reductions in conspiracy beliefs did not reliably translate into higher vaccination intentions. The authors conclude that these limited effects may be attributable to a failure to account for individuals' prior beliefs, emotional responses, and personal experiences.

These findings highlight the need for more personalized and context-sensitive interventions that target the psychological mechanisms underlying conspiracy beliefs

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

rather than relying solely on factual correction (Steinert et al., 2022; Taubert et al., 2024).

Large language models (LLMs) provide access to vast amounts of information and enable highly personalized interactions, creating new opportunities for information dissemination, education, and public engagement (e.g., OpenAI, 2024). At the same time, their capacity to generate fluent and persuasive text raises concerns about the amplification of misinformation and the manipulation of public opinion.

Recent research suggests, however, that LLMs may also play a constructive role in countering disinformation by supporting fact-checking, identifying misleading narratives, and generating corrective or prebunking messages tailored to individual users (Costello et al., 2024; Costello et al., 2025).

Costello et al. (2024) investigated whether conspiracy beliefs can be reduced through personalized, evidence-based dialogue with an LLM. The authors hypothesized that prior debunking failures may stem from interventions lacking sufficient depth and personalization. Across two large-scale experiments involving more than 2,000 U.S. participants, individuals engaged in short, tailored conversations with an AI model (GPT-4 Turbo) that directly addressed the specific evidence they cited in support of their preferred conspiracy theory. Compared with control conditions, these AI-led dialogues produced an average reduction in conspiracy belief of approximately 20%. Importantly, these effects persisted for at least two months and were observed across a wide range of conspiracy theories, including those related to COVID-19, political events, and long-standing historical claims. These findings suggest that many conspiracy beliefs may be more responsive to compelling, personalized counterevidence than previously assumed (Lewandowsky et al., 2013; Sunstein & Vermeule, 2009).

Building on these results, Costello et al. (2025) conducted a preregistered experiment with 1,297 conspiracy-believing participants to isolate the psychological mechanisms underlying this effect. When the AI was instructed to persuade participants without presenting factual counterevidence (e.g., by highlighting harms or relying on narrative examples), the debunking effect was largely eliminated or even reversed. In contrast, conditions that provided factual information without explicit persuasive intent were as effective as the baseline intervention. These findings align with emerging evidence that conspiracy beliefs are responsive to high-quality counterarguments when they directly address believers' specific claims (Altay et al., 2022; Costello et al., 2024).

Recent experimental work demonstrates that large language models can reduce conspiracy beliefs when they engage users in personalized, evidence-based dialogue that directly addresses the specific claims underlying those beliefs (Costello et al., 2024). Importantly, subsequent research indicates that these effects depend critically on the provision of high-quality factual counterevidence rather than on persuasive intent or emotional framing (Costello et al., 2025). While these findings suggest that conspiracy beliefs may be responsive to tailored rational engagement, existing evidence is largely limited to U.S. samples.

Moreover, their design allowed participants to choose any conspiracy belief they held. As a result, many of the beliefs targeted in their intervention may have differed substantially in strength, affective intensity, and motivational significance. While this approach has clear ecological advantages, it leaves open the question of whether the observed effects generalize to specific categories of conspiracy beliefs that may be especially strong or emotionally laden. Vaccine conspiracy beliefs are a particularly important case in this respect, because they are often emotionally charged, health-relevant, and connected to distrust of medical and governmental institutions. Consequently, vaccine conspiracies may be more affectively charged and more resistant to factual correction than many other conspiracy beliefs.

It also remains unclear whether different conversational styles—particularly those varying in their reliance on rational versus emotional arguments—produce differential effects on vaccine-related conspiracy beliefs in contexts characterized by high levels of vaccine hesitancy.

The present study addresses these gaps by examining the effects of LLM-mediated dialogue on vaccine conspiracy beliefs among Bulgarian citizens, systematically varying the conversational style of the interaction. Using a procedure adapted from Costello et al. (2024) and Costello et al. (2025) in the present study the focus is on vaccine conspiracies thus providing a more stringent test of whether AI-based dialogue can reduce beliefs that are highly emotional and personally relevant.

Goals and Hypotheses

The primary goal of this study is to test whether interaction with a LLM can reduce vaccine conspiracy beliefs in a non-U.S. population, thereby extending prior findings by Costello et al. (2024, 2025) to a new cultural and substantive context. A secondary goal is to examine whether the effectiveness of LLM-based debunking depends on the conversational style of the interaction. Specifically, the study compares three LLM dialogue conditions – a *free* argumentation style that addresses participants' conspiracy-related claims; a *rational*, evidence-based argumentation style that directly addresses participants' conspiracy-related claims; an *emotional* argumentation style that uses emotional appeals, empathy, concern, without systematic factual rebuttal. There is also a *control* conversational style in which participants interact with the LLM on a topic unrelated to vaccines.

Based on recent evidence demonstrating that large language models can reduce conspiracy beliefs when they provide personalized, evidence-based counterarguments (Costello et al., 2024), we hypothesize that participants who engage in LLM-mediated conversations directly addressing vaccine conspiracy beliefs will exhibit a reduction in belief endorsement relative to participants in a control condition involving an unrelated topic. This prediction extends prior findings to a non-U.S. context and to vaccine-specific conspiracy beliefs.

Furthermore, consistent with evidence that the effectiveness of AI-based debunking depends critically on the provision of high-quality factual counterevidence rather than persuasive intent (Costello et al., 2025), we expect belief change to vary as a function of argumentation style used by the LLM. Specifically, conversations grounded in rational, evidence-based argumentation are expected to produce greater reductions in conspiracy belief endorsement than conversations relying primarily on emotional appeals or unconstrained persuasion.

Finally, drawing on findings that emotionally framed interventions without direct engagement with evidence may be ineffective or counterproductive (Costello et al., 2025), we anticipate that emotionally framed LLM conversations will not reliably reduce vaccine conspiracy beliefs and may yield effects comparable to those observed in the free-form or control conditions.

The motivation for this paper stems from the central role that debates about vaccines have played in the aftermath of the COVID-19 pandemic. In this context, understanding vaccine-related conspiracy beliefs and identifying effective strategies for reducing them represent an important challenge. This task is particularly difficult, as evidence suggests that conspiracy beliefs concerning highly salient and socially important topics are especially resistant to change (Costello et al., 2024).

Method

Procedure and Design

The study consisted of two parts. In the first part, participants completed an online questionnaire (Google Forms) assessing demographics and individual differences, including conspiracy mentality (CMQ; Bruder et al., 2013), cognitive reflection (CRT; Frederick, 2005), and vaccine conspiracy beliefs (VCBS; Shapiro et al., 2016). A subset of participants also completed a single-item measure of vaccine-related conspiracy belief. All responses were provided anonymously and at participants' own pace. Data from the questionnaires are not reported in this paper.

The second (main) part involved an interactive conversation with ChatGPT-4o via a custom Zapier-based interface (creativity parameter = 0.7). At the beginning of the interaction, participants responded to two open-ended questions assessing their views on vaccine-related conspiracy theories. First, they were asked to describe any vaccine conspiracy theory they found convincing and explain why:

> There is a lot of debate about vaccines. Some people believe that information about vaccines—such as information about their safety or effectiveness—is deliberately manipulated or concealed by government institutions, pharmaceutical companies, and scientists. Others believe that vaccinating children is harmful and that this harm is being covered up, or that vaccination programs are used as a tool for population control. Are there any theories of this kind that you find particularly credible or convincing? Please describe one of them and explain why.

Participants were then asked to elaborate on the sources of their beliefs, including perceived evidence, events, personal experiences, or information sources:

> Would you be willing to share more about what makes you think this way? For example, are there specific evidence, events, sources of information, or personal experiences that have influenced your perspective?

Responses were provided in free-text format and combined into a single text reflecting each participant's expressed belief (opinion). This composite response was used for frequency analyses and to generate a one-sentence summary of each participant's position (summary) produced by ChatGPT. Participants then rated the perceived truth of this summary by typing a number from 0 to 100 (0 = definitely false, 50 = uncertain, 100 = definitely true), which served as the pre-conversation measure of vaccine-related conspiracy belief.

For the conversation with ChatGPT study used a between-subjects design with four experimental conditions that varied the conversational instructions provided to ChatGPT on what type of argumentation style to be used. Participants engaged in three conversational turns during which the chatbot attempted to challenge participants' vaccine conspiracy beliefs, except in the control condition.

In the *free argumentation condition*, ChatGPT was instructed to persuade participants without using any specified argumentative strategy. In the *rational argumentation condition*, the chatbot was instructed to rely exclusively on rational reasoning, scientific evidence, statistical data, and expert explanations, while avoiding emotional appeals, personal stories, or slogans. In the *emotional argumentation condition*, ChatGPT was instructed to rely on emotional appeals and social cues, using personal stories and short catchphrases while avoiding scientific evidence, statistics, or expert explanations. In the *control condition*, ChatGPT engaged participants in a neutral conversation about pets.

After the interaction, participants again rated the perceived truth of the same belief summary on the same 0–100 scale, providing the post-conversation measure. The primary dependent variable was the change in belief ratings from *before* conversation to *after* conversation, allowing assessment of the effects of conversational style on vaccine conspiracy beliefs.

Subsequently, participants completed a single-item measure of vaccine conspiracy belief and several items assessing behavioral intentions (e.g., willingness to receive booster vaccinations or seek further information; using 1–5 scale and Yes/No answers). At the end of the study, all participants were provided with a link to an official Bulgarian vaccine information website.

Participants

Participants were recruited via university research pools (for course credits) and social media. Of 438 adults who completed the initial online survey, 336 also completed the LLM interaction and were included in the analyses. Participants ranged in age from 18 to 76 years (M = 28.7, SD

= 12.7); 275 identified as women, 59 as men, and 2 did not report gender. The sample included 260 students (primarily psychology) and 76 non-students.

The study was approved by an ethics committee, conducted in accordance with the Declaration of Helsinki, and all participants provided informed consent. Data were collected anonymously.

Results

Frequency of Conspiratorial Beliefs about Vaccines

Participants' free-text responses to the two open-ended questions were combined into a single text reflecting their expressed opinion regarding vaccine-related conspiracy beliefs (opinion). The opinions were classified for the presence of conspiratorial beliefs about vaccines using an automated coding procedure implemented via a Zapier chatbot powered by ChatGPT-4o with creativity set to 0. The model was presented with a list of 20 possible conspiratorial claims about vaccines. The model classified whether the response contained any of the conspiratorial claims and whether the participant expressed belief in such claims (yes vs. no). Ratings about which conspiratorial claims are expressed is not presented here due to the lack of space. In the paper we present only the classification of the opinions as conspiratorial or not. A human coder independently evaluated a random subset of 50 responses if they present belief in vaccine conspiracies or not. Interrater agreement was high (Cohen's $\kappa = .776$). Given this level of agreement, ChatGPT-generated classifications were used in subsequent analyses.

Based on the coding of participants' opinions (open-ended responses), 124 out of 336 participants (36.9%) expressed conspiratorial beliefs related to vaccines, whereas 212 participants (63.1%) did not.

The high frequency of conspiratorial beliefs in the open-ended responses is notable: more than one-third of participants expressed at least one vaccine-related conspiracy belief. This suggests that vaccine-related conspiracy beliefs were not marginal in the sample, but were relatively widespread among participants.

Effect of the Conversation with a LLM on Changing Vaccine Conspiracy Beliefs

As explained above, ChatGPT generated a one-sentence summary of each participant's opinion (summary). These one-sentence summaries generated by ChatGPT were coded for the presence of conspiratorial beliefs by the authors. 127 out of 336 summaries (37.8%) expressed conspiratorial beliefs, whereas 209 summaries (62.2%) did not.

The classifications obtained from the two coding procedures were compared. For the subsequent analysis we included only these participants (110) for which both the original open-ended responses and the corresponding one-sentence summaries were classified as expressing vaccine-related conspiratorial beliefs.

Participants rated their belief in the summarized vaccine conspiracy before the conversation with the LLM and after the conversation on a 0-100 scale (0 – definitely false to 100 – definitely true).

An additional requirement was to include only participants who initially rated the summarized statement as at least 50 on the belief scale (108 participants). Six participants had missing data for the second rating. As a result, 102 participants were included in the analysis.

We explored if there is a change in the belief ratings after the conversation and if this change depends on the type of the arguments used by the model.

Mean *belief ratings* before and after the conversation for each experimental condition are presented in Table 1.

Table 1: Belief ratings *before* and *after* the conversation for each experimental condition.

Condition	N	Before conversation M (SD)	After conversation M (SD)	*p*
Free	25	81.1 (13.5)	80.4 (20.8)	.843
Rational	24	80.9 (14.5)	74.1 (20.9)	.006
Emotional	28	83.6 (14.6)	85.6 (15.1)	.124
Control	25	84.3 (13.4)	84.6 (13.2)	.832

Belief ratings were analyzed in a Repeated measures ANOVA with *rating time* (*before* conversation vs. *after* conversation) as a within-subjects factor and the *experimental condition* (*free* vs. *rational* vs. *emotional* vs. *control*) as a between-subjects factor. The analysis revealed an interaction between rating time and experimental condition ($F(3, 98) = 3.05$, $p = .032$, $\eta_p^2 = .085$), see Figure 1. Both main effects were not statistically significant.

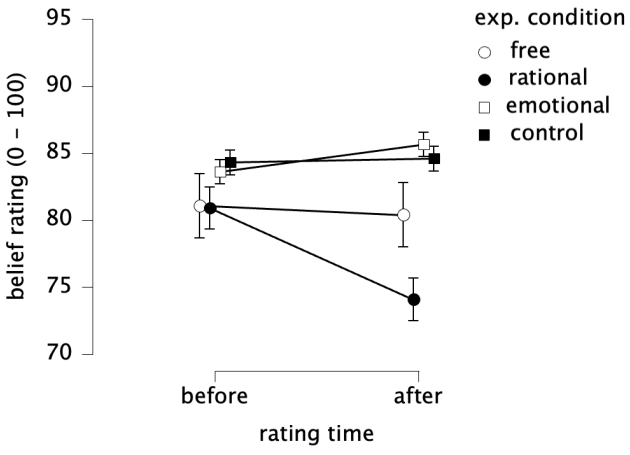

Figure 1: Belief ratings before and after the conversation for each experimental condition.

Simple main effects analyses of the effect of *rating time* were conducted separately for each experimental condition. No significant effects of *rating time* were found in the *free* argumentation condition ($F(1, 98) = 0.04$, $p = .843$), the *emotional* argumentation condition ($F(1, 98) = 2.52$, $p = .124$), or the *control* condition ($F(1, 98) = 0.05$, $p = .832$). A significant effect of rating time was observed only in the *rational* argumentation condition ($F(1, 98) = 9.36$, $p = .006$). – mean ratings decreased from $M = 80.9$ ($SD = 14.9$) before the conversation to M = *74.1* ($SD = 20.9$).

Next, the analysis of the changes on an individual level is conducted. The change in the *belief ratings* before conversation and after conversation is coded as decrease (the belief rating after conversation is lower), no change (the same rating is provided), or increase (the belief rating after conversation is higher). Data for each experimental condition is presented in Table 2.

Table 2: Number of individual-level changes in belief ratings from before conversation to after conversation for each experimental condition.

Condition	Decrease	No change	Increase
Free	2	19	4
Rational	8	15	1
Emotional	1	24	3
Control	1	22	2

In the *free argumentation condition* the average change was -0.68 points on the belief rating scale. At the individual level, belief ratings decreased for 2 participants (-75 points and -2 points on the belief rating scale), increased for 4 participants (5 points for 2 participants and 25 points for 2 participants on the belief rating scale), and remained unchanged for 19 participants (Figure 2).

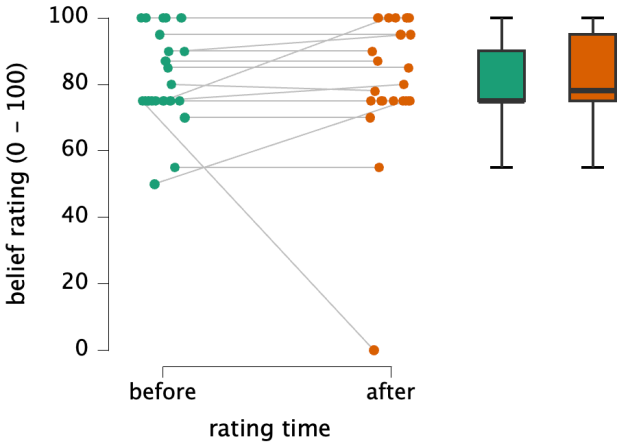

Figure 2: Belief ratings *before* and *after* the conversation for the *free argumentation condition*.

In the *rational argumentation condition* the average change was -6.83 points on the belief rating scale (as presented above, the effect of rating time was statistically significant). At the individual level, 8 participants showed a decrease in belief ratings (-25 points on the belief rating scale for 6 participants, -10 points for 1 participant, -5 points for 1 participant), 1 participant showed an increase (1 point), and 15 participants showed no change (see Figure 3).

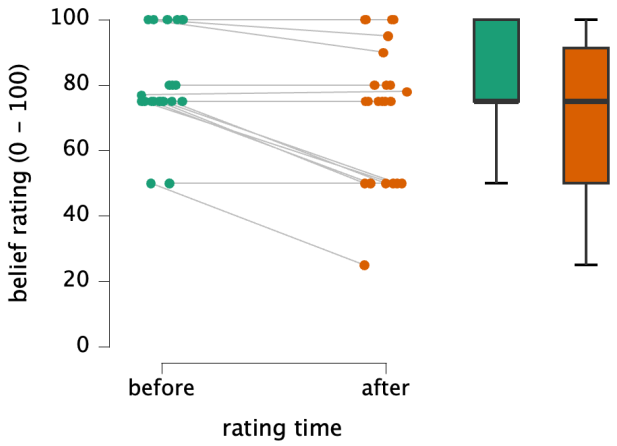

Figure 3: Belief ratings *before* and *after* the conversation for the *rational argumentation condition*.

In the *emotional argumentation condition* the average change was 2.03 points on the belief rating scale. At the individual level, belief ratings decreased for 1 participant (-3 points), increased for 3 participants (10, 25, 25 points), and remained unchanged for 24 participants (see Figure 4).

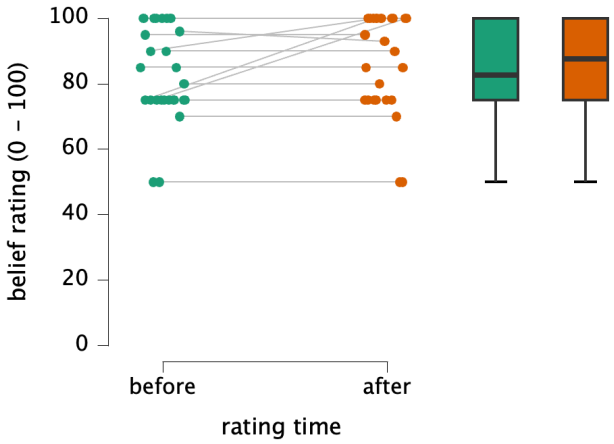

Figure 4: Belief ratings *before* and *after* the conversation for the *emotional argumentation condition*.

In the control condition the average change was 0.28 points on the belief rating scale. Individual-level analyses showed a decrease in belief ratings for 1 participant (-20 points), an increase for 2 participants (2 and 25 points), and no change for 22 participants (Figure 5).

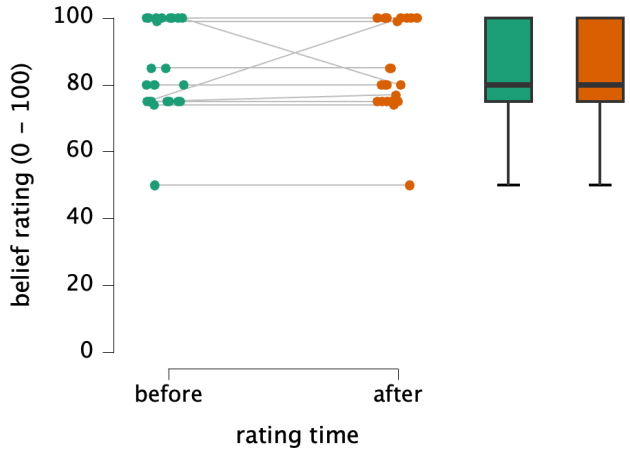

Figure 5: Belief ratings *before* and *after* the conversation for the *control condition*.

Vaccine-Related Intentions

Conspiracy supporting participants' vaccine-related intentions were assessed using three items measuring their willingness to receive a booster vaccination if contacted by their personal physician, receive recommended vaccinations when traveling to an exotic destination, and seek additional information regarding the necessity of booster vaccinations. Responses were provided on a 5-point Likert-type scale, ranging from 1 (no) to 5 (yes). A composite *intention score* was calculated by averaging responses across the three items, with higher values indicating greater willingness to engage in vaccine-related behaviors. Mean *intention scores* are presented in Table 3.

The intention score was analyzed in ANOVA with *experimental condition* (*free* vs. *rational* vs. *emotional* vs. *control*) as a between-subjects factor. The main effect of experimental condition on vaccine-related intentions did not reach statistical significance ($F(3, 97) = 1.07$, $p = .37$, $\eta_p^2 = .032$).

Table 3: Vaccine-related intentions for each experimental condition.

Experimental condition	*N*	*M*	*SD*
Free	25	3.6	1.2
Rational	24	3.8	1.0
Emotional	27	3.3	1.2
Control	25	3.3	1.2

The same analysis was conducted also for all participants (supportive or not of vaccine conspiracies). Again, the main effect of experimental condition on vaccine-related intentions did not reach statistical significance ($F(3, 314) = 1.97$, $p = .119$, $\eta_p^2 = .018$).

Discussion

The present study examined whether conversational interactions with a large language model can influence conspiratorial beliefs about vaccines and vaccine-related intentions, and whether such effects depend on argumentation style. Whereas previous studies examined participants' freely chosen conspiracy beliefs, the present study focused specifically on vaccine-related conspiracy beliefs among Bulgarian participants, a domain characterized by high public relevance, emotional salience, and elevated vaccine hesitancy.

The relatively high proportion of vaccine conspiaracies in the free responses (almost 37% of the participants) indicates that vaccine-related conspiratorial beliefs were common in the sample, underscoring the relevance of examining interventions aimed at reducing such beliefs.

Using an experimental manipulation of conversational strategies and pre–post conversational belief ratings, we found that rational, evidence-based argumentation was the only conversational approach associated with a reduction in vaccine conspiracy beliefs. No systematic change was observed in the free, emotional, or control conditions. These findings suggest that simply engaging in conversation, or appealing to emotions or social influence, is insufficient to reduce conspiratorial beliefs. Instead, belief change appears to require explicit engagement with evidence and logical reasoning, consistent with accounts emphasizing epistemic evaluation in belief revision.

In contrast to belief change, vaccine-related intentions were not significantly affected by experimental condition. This dissociation between belief endorsement and behavioral intentions suggests that reducing conspiratorial beliefs may not be sufficient to influence downstream health-related intentions, at least in the short term.

Methodologically, the study demonstrates the utility of LLMs as both experimental agents and analytic tools, supported by high agreement between automated and human coding. The use of open-ended elicitation enhanced ecological validity, while individual-level change analyses provided a nuanced view of intervention effects.

A limitation of the present study is the relatively small sample size that remained for the main analyses. Another limitation concerns the timing and format of the vaccine conspiracy belief ratings. Participants rated the perceived truth of the same summary before and after a relatively short conversation, which may have made the purpose of the study more transparent and introduced demand characteristics. In addition, participants may have remembered their initial rating, which could have influenced their post-conversation response. Future studies should address this issue by using improved measurement procedures.

Despite these limitations, the findings highlight both the potential and the limits of conversational AI in addressing misinformation. Rational argumentation can reduce conspiratorial belief endorsement, but it does not automatically translate into behavioral intentions.

Acknowledgments

The authors acknowledge the use of ChatGPT (OpenAI, 2024) as part of the experimental procedure and data gathering and processing.

The authors gratefully acknowledge Ivan Temelakiev for helpful discussions and assistance in setting up the ChatGPT chatbot, and Ilina Mutafchieva for technical assistance with running the study via Zapier.

References

Altay, S., Hacquin, A.-S., & Mercier, H. (2022). Why do so few people share fake news? It hurts their reputation. *New Media & Society*, *24*(6), 1303–1324. https://doi.org/10.1177/1461444820969893

Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). Measuring individual differences in generic beliefs in conspiracy theories across cultures: Conspiracy Mentality Questionnaire. *Frontiers in Psychology*, *4*, Article 225. https://doi.org/10.3389/fpsyg.2013.00225

Costello, T. H., Pennycook, G., Rand, D. G., & Gervais, W. M. (2024). Durably reducing conspiracy beliefs through dialog with AI. *Science*, *385*(6708), 145–150. https://doi.org/10.1126/science.adk8158

Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, *385*(6714), eadq1814. https://doi.org/10.1126/science.adq1814

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Jolley, D., & Douglas, K. M. (2014). The effects of anti-vaccine conspiracy theories on vaccination intentions. *PLoS ONE*, *9*(2), e89177. https://doi.org/10.1371/journal.pone.0089177

Lewandowsky, S., Oberauer, K., & Gignac, G. E. (2013). NASA faked the moon landing—Therefore, (climate) science is a hoax: An anatomy of the motivated rejection of science. *Psychological Science*, *24*(5), 622–633. https://doi.org/10.1177/0956797612457686

OECD/European Commission (2024), *Health at a Glance: Europe 2024: State of Health in the EU Cycle*, OECD Publishing, Paris, https://doi.org/10.1787/b3704e14-en

OpenAI (2024). ChatGPT (GPT-4) [Large language model]. https://chat.openai.com/

Our World in Data. (2026). *Share of people who completed the initial COVID-19 vaccination protocol* [Data set]. Retrieved May 5, 2026, from https://ourworldindata.org/grapher/share-people-fully-vaccinated-covid

Shapiro, G. K., Holding, A., Perez, S., Amsalem, D., & Rosberger, Z. (2016). The vaccine conspiracy beliefs scale: Development and validation of a measure of conspiracy beliefs about vaccination. *PLoS ONE*, *11*(10), e0164786. https://doi.org/10.1371/journal.pone.0164786

Steinert, J. I., Sternberg, H., Prince, H., Fasolo, B., Galizzi, M. M., Büthe, T., & Veltri, G. A. (2022). COVID-19 vaccine hesitancy in eight European countries: Prevalence, determinants, and heterogeneity. *Science Advances*, *8*(17), eabm9825. https://doi.org/10.1126/sciadv.abm9825

Sunstein, C. R., & Vermeule, A. (2009). Conspiracy theories: Causes and cures. *Journal of Political Philosophy*, *17*(2), 202–227. https://doi.org/10.1111/j.1467-9760.2008.00325.x

Taubert, F., Meyer-Hoeven, G., Schmid, P., Gerdes, P., & Betsch, C. (2024). Conspiracy narratives and vaccine hesitancy: A scoping review of prevalence, impact, and interventions. *BMC Public Health*, *24*, 3325. https://doi.org/10.1186/s12889-024-20797-y

UNICEF Bulgaria, & Bulgarian Ministry of Health (2023). Attitudes toward the application of the Bulgarian childhood immunization calendar among parents of children aged 0–4 years (National representative study). *UNICEF Bulgaria*.