

Personalized Persuasion Through Conversational AI: Can DeepSeek Change Perceptions of Genetically Modified Foods in China?

Qi Xi ^{1,2} , Jing Zeng ³ , Zhanghao Li ⁴ , and Mike S. Schäfer ³ 

¹ Center for Computational Communication Research, Beijing Normal University, China

² School of National Safety and Emergency Management, Beijing Normal University, China

³ Department of Communication and Media Research, University of Zurich, Switzerland

⁴ School of Journalism and Communication, Guangzhou University, China

Correspondence: Zhanghao Li (lizhanghao.comm@gmail.com)

Submitted: 29 September 2025 **Accepted:** 5 January 2026 **Published:** 5 March 2026

Issue: This article is part of the issue “Exploring Engagement With Complex Information: Perspectives on Generative AI as an Information Intermediary” edited by Monika Taddicken (TU Braunschweig), Esther Greussing (TU Braunschweig), Evelyn Jonas (TU Braunschweig), Ayelet Baram-Tsabari (Technion–Israel Institute of Technology), and Inbal Klein-Avraham (Technion–Israel Institute of Technology), fully open access at <https://doi.org/10.17645/mac.i509>

Abstract

Conversational AI has become an influential intermediary in public communication. Emerging research on conversational AI highlights its potential to correct misconceptions and influence attitudes across domains. This study investigates the persuasive effects of personalized conversational AI, focusing on genetically modified foods in China. Employing a between-subjects factorial design, 813 participants engaged in dialogues with a DeepSeek-based chatbot. Participants were randomly assigned to one of four conditions, ranging from a non-personalized generic control to increasingly tailored approaches based on demographic information, risk perceptions, or a combination of both. Results indicate that while AI interactions significantly improve attitudes and willingness to consume genetically modified foods across all conditions, the additional persuasive effect of personalization was conditional. Only a personalization strategy combining demographics and risk perception yielded greater persuasive effects than the control, primarily among participants with positive risk perceptions. Furthermore, moderation analyses revealed a divergence in individual differences: Among participants with negative risk perceptions, while greater prior experience with AI and higher trust in science decreased the persuasive effects, higher AI knowledge facilitated greater attitude gains.

Keywords

AI persuasion; conversational AI; generative AI; genetically modified foods; science communication

1. Introduction

Powered by large language models (LLMs), today's conversational AI systems like ChatGPT and DeepSeek have emerged as new intermediaries in public communication, bridging traditional information sources with interactive agents that can adapt to users' needs (Schäfer, 2023). For example, these systems have the capacity to tailor explanations to users' queries and knowledge levels, potentially improving understanding of complex scientific topics (C. Chen & Shu, 2024; Yan et al., 2024). Early evidence also highlights the persuasive potential of conversational AI across various domains. AI-mediated dialogues have been used to debunk misinformation and even reduce belief in conspiracy theories (Costello et al., 2024). Beyond improving knowledge or correcting misconceptions, conversational AI has shown persuasive potential that extends to influencing attitudes and driving behavior change. For instance, AI-generated messages have shown promise in encouraging positive behaviors in health and environmental contexts (Matz et al., 2024; Xia et al., 2024).

This persuasive potential is closely tied to the distinct communication affordances of conversational AI. On the one hand, unlike one-way media disseminations, conversational AI enables dialogic and personalized interactions, allowing users to ask questions, receive tailored responses, and engage in back-and-forth discussions. On the other hand, these dialogues are not one-size-fits-all solutions; they are personalized. Personalization—the adaptation of messages based on user characteristics—is widely regarded as a potent strategy in persuasion (Matz et al., 2024; Teeny & Matz, 2024). Personalized persuasion through AI could involve customizing content to align with a user's demographics, prior beliefs, or psychological attributes, thereby making the communication more relevant and compelling to a given individual.

Existing evidence indicates that messages generated by conversational AI can be more influential when tailored to participants' profiles than when they are generic one-size-fits-all messages (Karinshak et al., 2023; Matz et al., 2024). However, findings on AI-based personalization are controversial. Some scholars only found minimal differences between personalized and non-personalized AI-generated messages in political issues (Hackenburg & Margetts, 2024). And recent evidence from health communication suggests that personalization can sometimes even backfire: AI-generated corrections reinforced people's pseudoscientific beliefs, rather than reducing them (Lu, 2025). These results highlight that the effectiveness of personalization may depend on factors such as the issue at hand, the approach of personalization, or the participant's characteristics.

Situated within the field of science communication, the present study contributes to this growing body of research by using genetically modified foods (GMFs) in China as a case example to examine how conversational AI influences public attitudes toward complex scientific issues. In China, GMFs have long faced public skepticism and conspiratorial narratives despite the fact that scientific evidence supports their safety and benefits (Jiang & Fang, 2019; G. Wang et al., 2021). Public concerns stem not only from scientific uncertainty but also from social and political factors, including recurring food safety scandals, distrust in regulatory institutions, and geopolitical tensions (Cui & Shoemaker, 2018). These dynamics make GMFs in China an especially informative case for assessing conversational AI as a persuasive channel in a contested scientific domain, focusing on two outcomes: attitudes toward GMFs and willingness to consume. Therefore, employing online experiments, this study tests whether conversational AI can be persuasive around GMFs in China, and whether personalized messaging enhances this persuasive impact. Beyond science

communication, the findings of this study have important implications for the broader field of media and communication. As conversational AI systems are emerging as powerful communicative actors in contemporary society (Depounti & Natale, 2025), this study provides critical insights into how conversational AI-mediated interactions shape public attitudes and persuasion processes.

2. Literature Review

2.1. *Public Communication and Perception of GMFs*

Public and media communication plays a crucial role in shaping how societies engage with, understand, and perceive emerging technologies (Bucchi & Trench, 2021). This is particularly relevant for technologies that are complex, prone to misperceptions, and historically contested—such as GMFs. Despite their potential to address pressing agricultural and environmental challenges, public debates on GMFs remain controversial, frequently disconnected from scientific evidence, and vulnerable to distrust, misinformation, and conspiracy narratives (Ryan, 2014; Xu et al., 2021).

One factor contributing to challenges in public discussion is the inherent complexity of the technology itself. Concepts such as genome editing are highly technical and abstract, making it challenging to present information in an accessible and engaging way for non-specialist audiences. Communicators face the task of simplifying without oversimplifying—balancing accuracy with comprehensibility (Tonui et al., 2024). The complexity of genome editing makes it a compelling case for studying persuasion, as its technical opacity and the prevalence of polarized narratives create conditions in which communication strategies can critically shape public attitudes and behavioral intentions.

The dynamics outlined above are especially pronounced in China, where discussions of genome editing are further complicated by geopolitical concerns and nationalism. On the topic of GMFs in particular, conspiracy thinking has emerged as a persistent theme in public discourse in China (Jiang & Fang, 2019; Xu et al., 2021). For example, claims that genetically modified crops are a tool of US economic dominance, or that multinational corporations intentionally manipulate biotechnology to threaten China's food security, have gained traction among segments of the public (Cui & Shoemaker, 2018; Du et al., 2022). Such discourses underscore the necessity of effective communication interventions in China.

Public trust is another key factor shaping public perception of GMFs. Research shows that declining trust in regulatory bodies and scientific institutions central to biotechnology governance poses a significant barrier to constructive engagement with GMFs (Yuan et al., 2019). The decline in trust is often linked to perceived conflicts of interest, lack of transparency in decision-making, and assumptions that institutional actors prioritize commercial or political agendas over the public good (Marques et al., 2024).

In China, the dynamics around public trust are more complex. While trust in authorities and science tends to be higher than in many Western contexts (Cologna et al., 2025; Huang et al., 2021), skepticism nonetheless persists. Much of this skepticism stems from food safety concerns, past scandals involving contaminated products, and perceptions that official communication lacks openness and responsiveness to public concerns (Cui & Shoemaker, 2018). This creates a paradoxical environment in which relatively high baseline trust in institutions coexists with anxieties about GMFs specifically.

In addition, public engagement with and participation in technology debates can remain low in many contexts even when trust in science is relatively high. LLMs may help bridge this gap by providing accessible, interactive, and tailored exchanges; they can make complex issues more relatable and lower entry barriers (Schäfer, 2023). This suggests that beyond informing the public, LLMs hold potential to actively expand participation in science-related discussions in China (Savaget et al., 2018; Schäfer, 2023).

2.2. Conversational AI, Public Engagement With Science, and Persuasion

Building on the context outlined above, conversational AI can be a potent tool to foster public engagement with complex and contested technologies. By enabling dialogic and personalized interaction, conversational AI systems such as ChatGPT and DeepSeek create avenues for more engaging and dialogical science communication (Schäfer, 2023).

Conversational AI reshapes the way people seek and engage with scientific information, increasingly functioning as an alternative to traditional search engines (Zhou & Li, 2024). These systems provide interactive, on-demand explanations that support information seeking on science-related topics (Greussing et al., 2025). This shift is particularly relevant for biotechnology, where conversational AI can make complex technical knowledge more accessible and thereby support self-initiated science learning.

It is also important to note that scholars caution that conversational AI tools may generate information that is biased, misleading, or potentially harmful (Ta et al., 2025). In particular, current conversational AI systems are prone to the much-discussed “hallucination” problem, whereby they generate outputs that are factually incorrect or unverifiable while presenting them as accurate information (Lee et al., 2025; Shao, 2025). Similar concerns have been raised regarding AI-generated information on gene-editing-related scientific topics, with research showing that these systems may reproduce biases or inaccuracies (Burright & Al-khateeb, 2025).

At the same time, a growing body of scholarship has begun to examine the potential of conversational AI’s persuasive quality, such as in countering misinformation and correcting biases (C. Chen & Shu, 2024; Karinshak et al., 2023). For instance, conversational AI can be deployed to debunk myths, deliver accurate information in real time, and strengthen users’ ability to evaluate the credibility of scientific claims (K. Chen et al., 2024; Karinshak et al., 2023). Recent research has also investigated the design of conversational agents to reduce belief in conspiracy theories (Costello et al., 2024) and to influence public perceptions of contested science topics such as agricultural biotechnology (Ou et al., 2025). Although still an emerging area of study, prior evidence highlights the persuasive potential of conversational AI.

2.3. Personalization in AI Persuasion

In addition, scholars suggest that personalization may enhance AI’s persuasive capacity. A growing body of literature shows that conversational AI systems can already generate persuasive messages across a range of contexts, sometimes outperforming institutional or human baselines (Karinshak et al., 2023). In the context of AI persuasion, personalization typically involves tailoring message content to user characteristics such as demographics, prior beliefs, or other traits. By aligning communication with these attributes, personalization is expected to increase information acceptance and reduce resistance (Kaptein et al., 2015). Research also indicates that conversational AI systems can mitigate the egocentrism biases that often constrain human

communicators, allowing them to more accurately address users' specific needs (Matz et al., 2024). Moreover, conversational AI's dynamic learning and iterative optimization capabilities allow it to accumulate user data through logged chat history over time. This mechanism can support the development of detailed user profiles, potentially further enhancing both the precision and the persuasive impact of personalized content.

However, evidence on the effectiveness of personalization is mixed, and effects seem context-dependent. For instance, Hackenburg and Margetts's (2024) experiment on persuasion on political issues found that personalization based on sociodemographic data did not significantly outperform non-personalized messages. In contrast, other studies demonstrate that even coarse sociodemographic information can enhance conversational AI's persuasive performance in interactive debate settings (Liu et al., 2022; Salvi et al., 2025).

Together, these findings imply that personalization can matter, but its effectiveness depends on what and how to personalise. Some scholars argue that sociodemographic attributes are both behaviorally influential and practically available, making them especially relevant for real-world AI-mediated communication (Hackenburg & Margetts, 2024; Liu et al., 2022). Others contend that personalization should focus on deeper, psychologically meaningful constructs such as personality traits, moral foundations, or prior beliefs, which are more directly tied to cognitive and attitudinal change (Teeny & Matz, 2024). Building on these insights, we consider both "shallow" attributes (e.g., demographics) and "deeper" attributes as bases for personalization. In the context of GMF communication, deeper attributes include existing risk perceptions. As informed by existing evidence, health and environmental risks are central sources of public resistance to GMFs (Nelson, 2001), and individuals with limited knowledge are particularly susceptible to biases regarding these risks. Accordingly, risk perception of GMFs represents a critical factor in persuasion (W. Zhang et al., 2019). Drawing on these considerations, the study proposes the following hypotheses:

H1: Personalized persuasion is more effective than non-personalized persuasion in improving attitudes toward GMFs and willingness to consume GMFs.

H1a: Demographic-based personalization yields greater improvements than non-personalized messaging.

H1b: Belief-based personalization (tailored to risk perceptions) yields greater improvements than non-personalized messaging.

H1c: Combined demographic- and belief-based personalization yields greater improvements than non-personalized messaging.

Beyond message design, persuasion varies on contested issues. Experimental studies consistently show that persuasive interventions often generate heterogeneous treatment effects, with outcomes shaped by individual-level predispositions (Costello et al., 2024). To account for this, the present study considers two moderators: AI literacy and trust in science. AI literacy encompasses competencies such as knowledge of and experience with applying AI systems. Prior research suggests that AI literacy influences how individuals perceive the credibility, usefulness, and persuasiveness of AI systems (Shin, 2022; C. Wang et al., 2025).

Furthermore, in the context of GMF communication, trust in science is another important factor that can influence one's perception and acceptance of biotechnology (Master & Resnik, 2013; Priest et al., 2003). Evidence shows that greater levels of trust in science generate greater citizen support for biotechnology, including agricultural biotechnology (see review in Brossard & Nisbet, 2007).

Based on these insights, we propose the following hypotheses regarding participants' heterogeneity:

H2: The persuasive effects of the same personalization approach vary across audience subgroups.

H2a: Individuals with higher AI literacy respond more positively to personalized persuasion than those with lower AI literacy.

H2b: Individuals with greater trust in science respond more positively to personalized persuasion than those with lower trust.

3. Methods

3.1. Experimental Design and Participants

We conducted a between-subjects factorial experiment, including four conditions (Personalization: control, demographic only, risk perceptions only, demographic + risk perceptions). Participants were randomly assigned to one condition. In personalized groups, we embedded demographic information and/or pre-existing briefs towards GMFs collected from the pre-test in the system prompt to customize the responses.

We chose DeepSeek-V3 as the backend for our experiment due to its open-source nature in technical aspects and its capability for rapid and high-quality content generation responses, especially in the Chinese language (Wu et al., 2025). DeepSeek was also one of the best open-weight chat models when we conducted this study. To embed the DeepSeek chatting interface in our questionnaire, we built a LangChain pipeline to receive users' input and store the model's output. The system prompts used in the study are provided in Table A.1 in the online appendix (available at https://osf.io/8pz4y/overview?view_only=0f9c28c16d664ee9889abc41856eaa64). For participants to access both the questionnaire and the AI conversation, we built a web application via Streamlit. Detailed information about technical implementation can be found in the online appendix (Section A.2).

We recruited 813 participants through two panel providers in China. We combined Wjx and Credamo to recruit participants. While Wjx provides samples similar to the national benchmark, Credamo offers samples that include more young and highly educated individuals (Del Ponte et al., 2024). Each participant received a compensation of 15 RMB (approximately 2 USD) for completing the study. The sample characteristics were as follows: 52% female and 48% male. Age distribution showed 14% aged 18–24, 47% aged 25–34, 30% aged 35–44, 6% aged 45–55, and 2% over 55 years. Regarding education, 94% of participants had attained some level of college education. Geographically, 87% of participants resided in urban areas. This study has received ethical approval from the Institutional Review Board (IRB) of Beijing Normal University. The approval documentation is provided in the online appendix.

3.2. Procedure and Stimulus

The study consisted of three phases: pre-test, interaction, and post-test (see Figure 1). First, in the pre-test phase, participants provided demographic data covering age, gender, education, and region of residence. In accordance with our hypotheses listed in Section 2.3, we also measured participants' AI literacy and trust in science. The Chinese public's acceptance of GMFs is shaped by factors like historical risk events, information transparency, government trust, and health and environmental risk concerns (Jin et al., 2022). Thus, we view risk perception of GMFs as a key factor in persuasion. Finally, participants reported their baseline attitudes and consumption intentions regarding GMFs. When participants finished the pre-test, their characteristics were used in the system prompt for personalizing the outputs. In the control group, the system received no participant-level information and produced generic, non-personalized responses. In experimental group I, the conversational AI system was allowed to access participants' demographic information; in experimental group II, the AI system was given access to their risk perception; and in experimental group III, both types of information were provided.

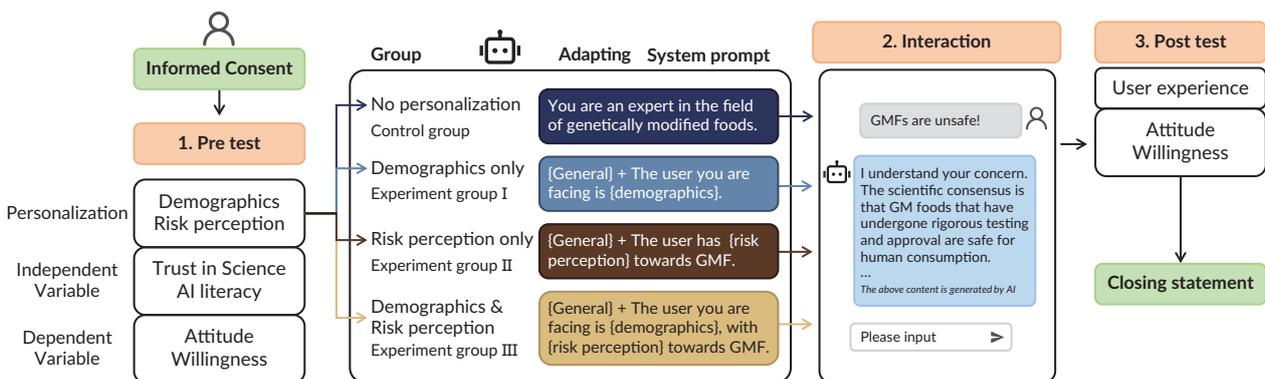


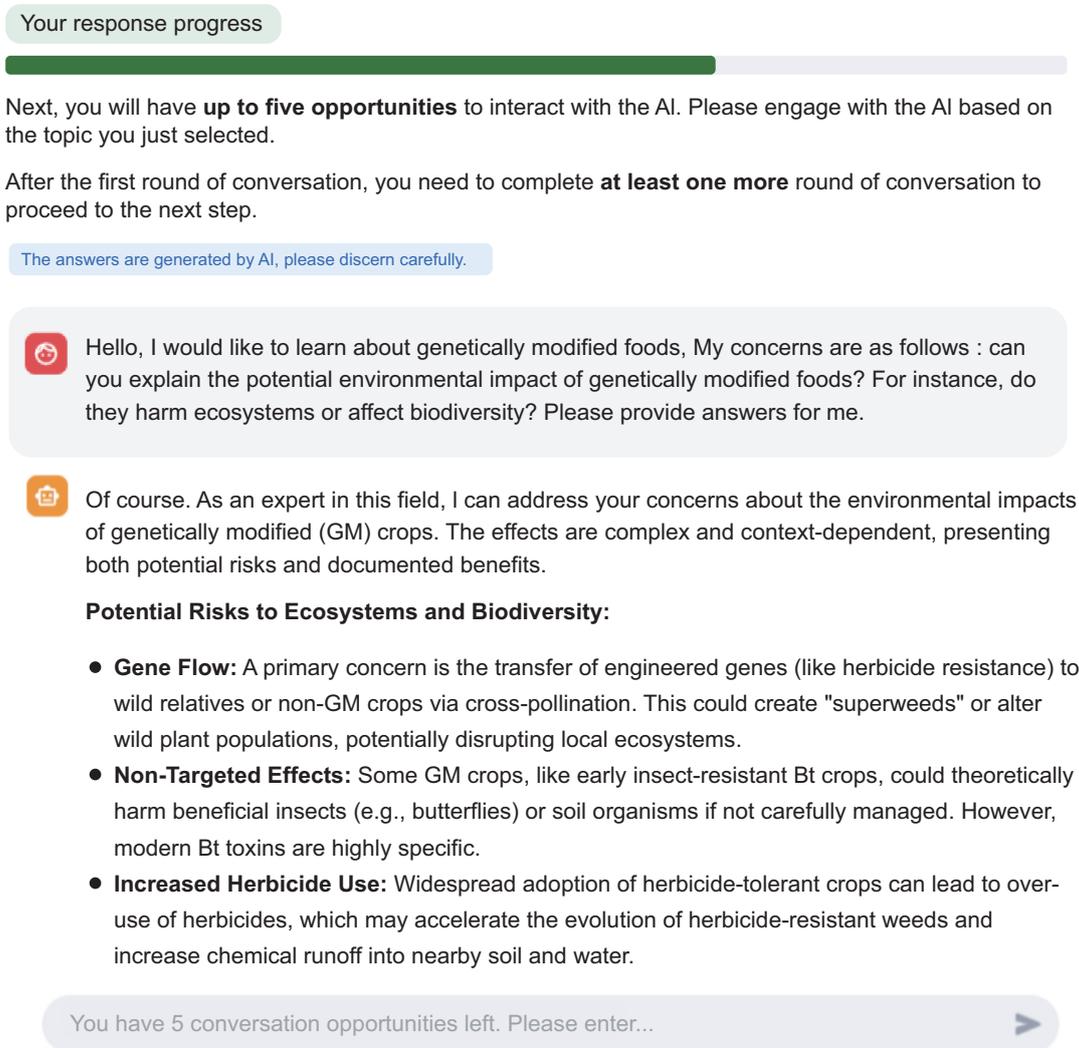
Figure 1. Experiment workflow.

Participants were randomly assigned to one of these groups. For demographics, there were no significant differences in gender ($\chi^2(3) = 4.22, p = 0.24$), education level ($\chi^2(6) = 5.45, p = 0.49$), or residence area ($\chi^2(3) = 2.89, p = 0.41$) between the four conditions. A one-way ANOVA showed no significant differences in age ($F(3,809) = 0.52, p = 0.67$). For covariates, a series of ANOVAs showed no significant differences in AI knowledge ($F(3,809) = 0.11, p = 0.95$), AI use ($F(3,809) = 0.54, p = 0.65$), trust in science ($F(3,809) = 0.20, p = 0.90$), or risk perceptions ($F(3,809) = 0.76, p = 0.52$). In pre-test measuring, ANOVAs showed no significant differences in attitude ($F(3,809) = 0.95, p = 0.42$) or willingness to consume GMFs ($F(3,809) = 1.06, p = 0.36$). Therefore, the randomization process can be considered effective, with no significant baseline differences among conditions.

The core experimental procedure—the interaction phase—involved two steps. First, the participants were asked to share their major concerns over GMFs across four dimensions: economic effects, ecological effects, potential health effects, and moral implications (Bawa & Anilakumar, 2013). Second, their responses were used as input for a controlled interaction with the conversational AI. Drawing on previous experimental designs (K. Chen et al., 2024; Costello et al., 2024; Hackenburg & Margetts, 2024), participants were allowed to interact with AI two to six times. That is, excluding the initial question expressing participants' concerns about GMFs, participants were required to ask one to five additional questions. The distribution of

interaction counts can be found in Table A.9 of the online appendix. The number of sessions had a non-significant difference ($F(3,809) = 1.23, p = 0.30$) across groups. In this phase, participants were informed at recruitment that each dialogue round should last at least 10 seconds, so as to encourage focused engagement and careful consideration of the AI's responses. This requirement was subsequently used by the panel provider as part of its quality control procedures for the recruited participants.

This requirement was subsequently checked and used to exclude participants from the final analysis. Personalization was implemented through system-level prompts that operated in the backend and were not visible to participants. Participants interacted with the AI through the same chat interface across all conditions and did not receive any additional instructions or prompts during the interaction. A compliance label was displayed across all conditions. Screenshots of this design are shown in Figure 2.



Your response progress

Next, you will have **up to five opportunities** to interact with the AI. Please engage with the AI based on the topic you just selected.

After the first round of conversation, you need to complete **at least one more** round of conversation to proceed to the next step.

The answers are generated by AI, please discern carefully.

Hello, I would like to learn about genetically modified foods, My concerns are as follows : can you explain the potential environmental impact of genetically modified foods? For instance, do they harm ecosystems or affect biodiversity? Please provide answers for me.

Of course. As an expert in this field, I can address your concerns about the environmental impacts of genetically modified (GM) crops. The effects are complex and context-dependent, presenting both potential risks and documented benefits.

Potential Risks to Ecosystems and Biodiversity:

- **Gene Flow:** A primary concern is the transfer of engineered genes (like herbicide resistance) to wild relatives or non-GM crops via cross-pollination. This could create "superweeds" or alter wild plant populations, potentially disrupting local ecosystems.
- **Non-Targeted Effects:** Some GM crops, like early insect-resistant Bt crops, could theoretically harm beneficial insects (e.g., butterflies) or soil organisms if not carefully managed. However, modern Bt toxins are highly specific.
- **Increased Herbicide Use:** Widespread adoption of herbicide-tolerant crops can lead to over-use of herbicides, which may accelerate the evolution of herbicide-resistant weeds and increase chemical runoff into nearby soil and water.

You have 5 conversation opportunities left. Please enter...

Figure 2. Screenshot of conversation interface. Note: translated from Chinese to English.

In the post-test phase, we first collected participants' feedback on their AI interactions during the experiment, as this helped us learn the extent to which participants chatted with the AI system. We then reassessed their attitudes towards GMFs, and their willingness to consume GMFs, using the same items as in the pre-test.

Qualified participants passed two attention checks, one in pre-test and one in post-test. Currently, there is no standardized manipulation check method for personalized experiments based on LLMs (Teeny & Matz, 2024). Previous studies have indicated that personalization generally leads to a better user experience (Alslaity et al., 2023); therefore, participants in the personalized experimental groups would be expected to enjoy a more positive experience when interacting with the AI. Accordingly, we analyzed the differences in subjective experience (user experience during the survey) between the control group and the aggregated set of three personalized groups (non-personalized vs. personalized: $F(1,811) = 4.50, p = 0.03$). Participants who took the personalization treatment reported more positive user experience compared to those in the control group, indicating that personalization led to differences in experience. This demonstrates that our manipulation was effective.

3.3. Measures

We measured AI literacy, trust in science, risk perceptions, and user experience using scales from existing studies. We used two simple 7-point Likert scales to score attitude towards GMFs and willingness to consume GMFs. The details of our scales can be accessed in the online appendix (Section A.3).

In regard to attitudes toward GMFs, participants completed four items assessing their attitudes toward GMFs. The included statements were issues that the Chinese public may concern (Jin et al., 2022), such as “GMFs are safe for the human body,” “GMFs have economic value,” “GMFs do not cause environmental damage,” and “I find GMFs acceptable from an ethical perspective” (for pre-test, $M = 3.77, SD = 0.35, \alpha = 0.92$; for post-test, $M = 4.62, SD = 0.31, \alpha = 0.89$).

In regard to willingness to consume GMFs, we used the question from Yuan et al. (2019) to measure willingness to purchase GMFs. Beyond purchasing, we also asked participants’ willingness to eat and share GMFs. A composite was created from the average of the three items (for pre-test, $M = 3.59, SD = 0.18, \alpha = 0.91$; for post-test, $M = 4.61, SD = 0.10, \alpha = 0.91$). This variable was designed to comprehensively examine the persuasive effect; the fitting and analysis with it as the dependent variable can be found in the online appendix (Section A.5).

To assess AI literacy, participants were asked to report their self-perceived AI literacy (Carolus et al., 2023) using a 7-point Likert scale (1 = *strongly disagree*, 7 = *strongly agree*). This measure captures participants’ subjective perceptions of their AI-related knowledge and experience. Specifically, the scale includes two dimensions: perceived AI knowledge, reflecting participants’ conceptual understanding of AI ($M = 5.37, SD = 0.22, \alpha = 0.71$), and perceived AI experience, reflecting their self-reported practical engagement with AI systems ($M = 5.65, SD = 0.11, \alpha = 0.77$).

To assess trust in science, participants responded to six items adapted from (Cologna et al., 2025) assessing trust in scientists on a 7-point Likert scale, from 1 = *strongly disagree* to 7 = *strongly agree*. Items included perceptions of scientists as honest, concerned for others’ wellbeing, ethical, eager to improve others’ lives, sincere, and considerate of others’ interests ($M = 5.05, SD = 0.13, \alpha = 0.87$).

In regard to risk perception of GMFs, we used five risk-related questions on a 7-point Likert scale (1 = *strongly disagree*, 7 = *strongly agree*) from Klimmt and Spreen (2025) to measure beliefs in risk perception

about GMFs. These items are reverse-coded. A composite was created from the average of the five items ($M = 3.46$, $SD = 0.09$, $\alpha = 0.92$). Inspired by G. Wang et al. (2021), we defined positive subgroups ($n = 257$) as participants whose risk perception was higher than 4, and the remaining were negative subgroups ($n = 556$).

Additionally, to gain an overview of users' experience with the interface—which can inform improvements to the experimental design in future studies—we also included questions on user experience. To do so, we employed measures from K. Chen et al. (2024), which include (a) ratings of the conversational AI system, (b) satisfaction with the dialogue, (c) learning experience with the AI system, and their intention to (d) continue the chat or (e) recommend the AI system to others. A composite was created from the average of the five items ($M = 5.67$, $SD = 0.12$, $\alpha = 0.76$).

3.4. Data Analysis Strategy

In this study, we analyzed data using Python and JASP. The statistical methods applied were paired t -test, chi-square test, analysis of variance (ANOVA), and linear mixed-effects model (LMM). For the LMM, we converted our data into long format. This format displays the data per measurement and separates it row by row; this effectively doubles the number of observations. Additionally, we adopted the fact-checking pipeline proposed by Lin et al. (2025) and Hackenburg et al. (2025) to assess the accuracy of the LLM outputs. In our experiments, 88% of the factual claims generated by the LLMs were traceable to credible and verifiable sources. Analysis procedures and results are available in the online appendix (Section A.6).

4. Results

4.1. Persuasive Impacts of Conversational AI

First, we conducted two paired t -tests to examine whether there were differences between the pre- and post-tests. The results indicated significant changes in both attitude ($t(812) = 26.70$, $p < 0.001$, Cohen's $d = 0.58$) and willingness ($t(812) = 26.57$, $p < 0.001$, Cohen's $d = 0.65$).

To determine the impact of persuasion, we fitted an LMM (model 1) to attitudes with whole samples ($n = 812$) first—see Table 1. The LMM showed that conversational AI successfully persuaded participants ($b = 2.615$, $SE = 0.281$, $p < 0.001$). In this model, age ($b = -0.013$, $SE = 0.004$, $p < 0.001$) and knowledge about AI ($b = -0.216$, $SE = 0.068$, $p < 0.001$) negatively predicted the attitude. AI experience ($b = 0.503$, $SE = 0.069$, $p < 0.001$), user experience ($b = 0.310$, $SE = 0.052$, $p < 0.001$), and trust in science ($b = 0.297$, $SE = 0.041$, $p < 0.001$) positively predicted the attitude towards GMFs. Compared with samples with positive risk perceptions, those with negative risk perceptions showed a more negative attitude ($b = -1.722$, $SE = 0.067$, $p < 0.001$).

To assess between-group persuasion impacts, we set interaction items between variables. For personalization, providing the LLM with both demographics and risk perceptions information resulted in a significant increase in impact increment compared to the control ($b = 0.193$, $SE = 0.090$, $p = 0.031$); it led to higher persuasion impacts, which supported H1c. However, neither demographics-only personalization ($b = 0.060$, $SE = 0.088$, $p = 0.494$) nor risk-perception-only personalization ($b = 0.107$, $SE = 0.088$, $p = 0.222$) produced a significant

gain over the control condition; therefore, H1a and H1b were not supported. Figure 3 illustrates the attitude changes of different interventions.

Table 1. LMMs results.

	model 1		model 1p		model 1n	
	b (SE)	p	b (SE)	p	b (SE)	p
(Intercept)	0.631 (0.365)	0.084	0.314 (0.568)	0.580	1.028 (0.438)	0.019
Individual differences						
Gender (male vs. female)	-0.119 (0.055)	0.029	-0.009 (0.057)	0.880	-0.216 (0.072)	0.003
Residence area (urban vs. rural)	-0.089 (0.083)	0.281	-0.020 (0.094)	0.830	-0.142 (0.105)	0.175
Education level (vs. undergraduate)						
Master's or higher	0.149 (0.080)	0.063	-0.019 (0.076)	0.803	0.267 (0.111)	0.016
Senior high or lower	-0.099 (0.122)	0.416	-0.540 (0.230)	0.019	-0.025 (0.142)	0.862
Age	-0.013 (0.004)	<0.001	0.001 (0.003)	0.850	-0.026 (0.005)	<0.001
AI knowledge	-0.216 (0.068)	0.001	0.293 (0.077)	<0.001	-0.524 (0.088)	<0.001
AI experience	0.503 (0.069)	<0.001	0.208 (0.095)	0.028	0.636 (0.086)	<0.001
User experience	0.310 (0.052)	<0.001	0.129 (0.074)	0.080	0.280 (0.061)	<0.001
Trust in science	0.297 (0.041)	<0.001	0.314 (0.061)	<0.001	0.150 (0.049)	0.002
Risk perception (vs. positive)	-1.722 (0.067)	<0.001	–	–	–	–
Group condition (vs. control)						
Demographic only	-0.034 (0.087)	0.700	-0.238 (0.091)	0.009	0.032 (0.114)	0.781
Risk perception only	-0.099 (0.088)	0.257	-0.163 (0.097)	0.094	-0.077 (0.111)	0.486
Demographic + Risk perception	-0.156 (0.089)	0.079	-0.176 (0.095)	0.063	-0.115 (0.116)	0.320
Treatment (post vs. pre)	2.615 (0.281)	<0.001	0.391 (0.623)	0.530	0.924 (0.346)	0.008
Treatment × individual differences						
Post × AI knowledge	0.053 (0.067)	0.426	-0.111 (0.083)	0.182	0.170 (0.083)	0.042
Post × AI experience	-0.254 (0.068)	<0.001	-0.113 (0.105)	0.281	-0.217 (0.081)	0.007
Post × User experience	0.064 (0.052)	0.214	0.320 (0.081)	<0.001	0.110 (0.058)	0.059
Post × Trust in science	-0.211 (0.039)	<0.001	-0.169 (0.068)	0.012	-0.029 (0.046)	0.535
Treatment × group interaction						
Post × Demographic only	0.060 (0.088)	0.494	0.197 (0.101)	0.051	0.050 (0.108)	0.644
Post × Risk perception only	0.107 (0.088)	0.222	0.166 (0.108)	0.126	0.030 (0.105)	0.777
Post × Demographic + Risk perception	0.193 (0.090)	0.031	0.263 (0.105)	0.012	0.157 (0.109)	0.152
Marginal R²	0.71		0.45		0.54	
Conditional R²	0.85		0.66		0.79	

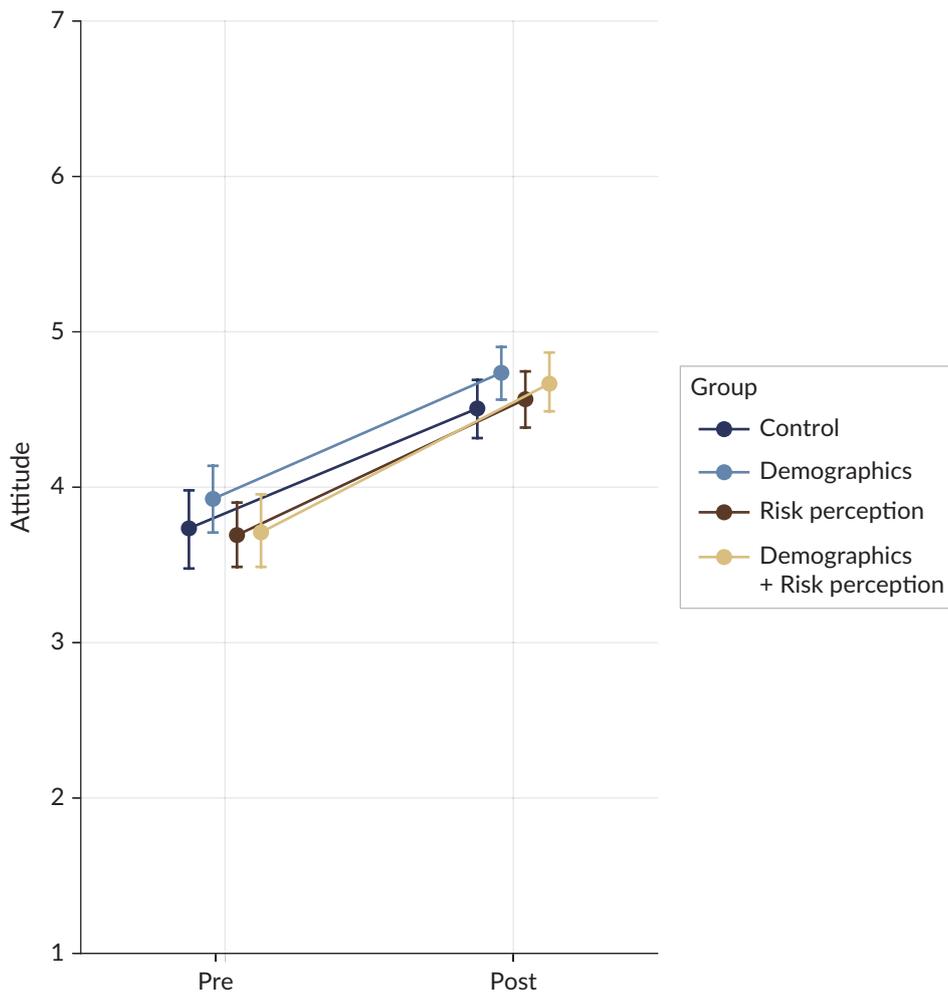


Figure 3. Persuasive impacts on attitude under different personalization approaches. Note: The error bars represent 95% confidence intervals.

In the interaction between covariates and treatment, we found that AI experience ($b = -0.254$, $SE = 0.068$, $p < 0.001$) and trust in science ($b = -0.211$, $SE = 0.039$, $p < 0.001$) both negatively influence persuasion outcomes, which rejected H2a and H2b. This indicates that respondents' experience with AI usage and their level of trust in science weaken the persuasive effect, leading them to express skepticism toward the related discussions.

4.2. The Heterogeneity of Persuasion Mechanisms Under Different Risk Perceptions

We divided the sample into two subgroups: negative risk perceptions ($n = 556$), referring to participants who overlooked the risks of GMFs, and positive risk perceptions ($n = 257$), referring to participants who held relatively objective opinions. Figure 4 illustrates the persuasion impact within these two subsamples. To avoid the complexity of introducing numerous interaction terms in a single model, we fitted an LMM separately for each subgroup. We analyzed the negative risk perceptions subgroup using model 1n, while we used model 1p to look at the positive risk perceptions subgroup, as detailed in Table 1.

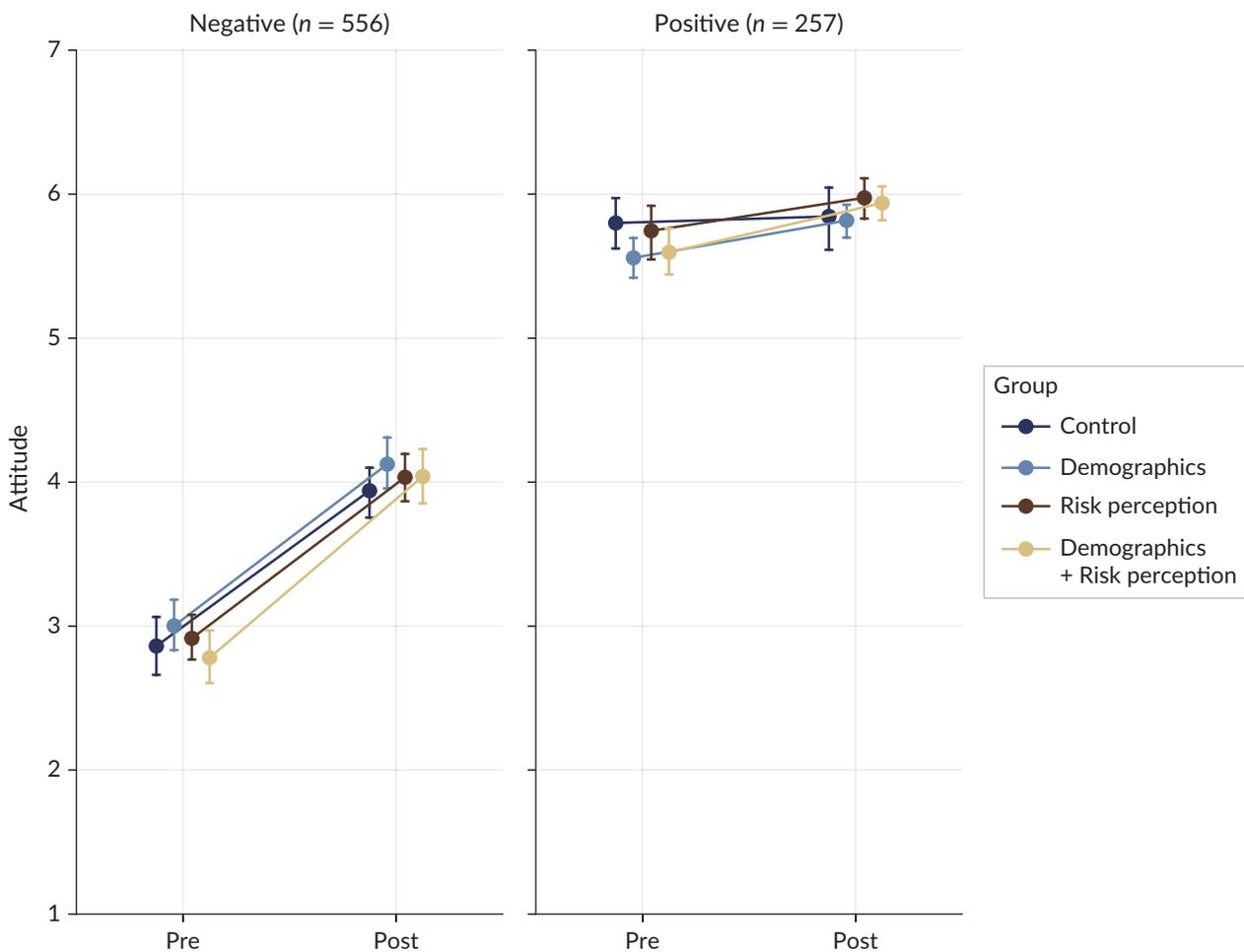


Figure 4. Persuasion impacts on attitude differed by risk perceptions. Note: The error bars represent 95% confidence intervals.

We analyzed the differences in persuasion impacts among three personalization methods across two subpopulations. In the negative risk perceptions subgroup, none of the personalized groups resulted in significant impact gains. This indicates that just chatting with AI already resulted in the maximum persuasion impact for those with negative risk perceptions. However, in the positive risk perceptions subgroup, providing both demographic and risk perceptions information led to significant improvements compared to control ($b = 0.263$, $SE = 0.105$, $p = 0.012$), while providing only demographic information was marginally significant ($b = 0.197$, $SE = 0.101$, $p = 0.051$). This indicates how H1c remains valid under the positive risk perception condition, but H1a is not supported in this case.

We also analyzed the differences in the effects of covariates in two subgroups. In the subgroup with positive risk perceptions, user experience improved the persuasive impacts ($b = 0.320$, $SE = 0.081$, $p < 0.001$), while trust in science weakened its impact ($b = -0.169$, $SE = 0.068$, $p = 0.012$). One possible explanation is that higher levels of trust in science lead to greater public engagement with science-related information, which may translate into greater skepticism towards AI-mediated persuasion (Schäfer et al., 2018). In the subgroup with negative risk perceptions, AI literacy played an important role in persuasion. Knowledge about AI increased persuasive outcomes ($b = 0.170$, $SE = 0.083$, $p = 0.042$), while frequent AI use reduced outcomes ($b = -0.217$, $SE = 0.081$, $p = 0.007$). This pattern suggests that conceptual understanding of AI may foster baseline trust

(Yang & Wibowo, 2022), whereas repeated exposure and hands-on experience can heighten sensitivity to AI limitations or failures, thereby dampening persuasion (Jones-Jang & Park, 2023; Lucas et al., 2024). Together, these results inform or prompt the more detailed theoretical discussion in Section 5.3.

5. Discussion and Conclusion

Our study contributes to the growing field of AI-mediated persuasion, with a focus on complex and controversial scientific topics like GMFs in China. We demonstrate that conversational AI, even when used in a simple, short-term interaction, possesses a significant persuasive capacity and can change public attitudes and willingness to consume GMFs. More importantly, our findings move beyond the question of whether AI persuasion works; rather, they address the more critical question of when and how it is most effective, echoing the call by Teeny and Matz (2024).

5.1. The Conditional Effectiveness of Personalization

The present study contributes to scholarly research on AI persuasion by showing that personalization can strengthen persuasion—but only when shallow audience cues (demographics) are combined with deeper, issue-proximal cues (risk perceptions). Single-cue personalization (demographics-only or risk-perception-only) did not outperform the control. This result complements recent arguments that research must shift from asking if conversational AI enhances personalization to identifying the conditions under which it does so (Teeny & Matz, 2024) and clarifies mixed evidence on microtargeting with LLMs (Hackenburg & Margetts, 2024; Matz et al., 2024). Conceptually, pairing demographic context with risk-focused concerns likely increases diagnostic relevance and reduces counter-arguing—mechanisms consistent with work on persuasion profiling in adaptive systems (Kaptein et al., 2015)—whereas either cue in isolation lacks sufficient specificity to generate incremental lift over generic content (Hackenburg & Margetts, 2024).

5.2. Individual Differences and Persuasive Heterogeneity

Relative to the control condition, none of the personalization approaches increased persuasion among participants with high risk perceptions of GMFs. Among participants with lower perceived risk, combined personalization yielded the largest gains, while single-cue tailoring did not reliably exceed the control. One factor that may help explain this pattern is the complex sociopolitical context surrounding GMF debates in China. As discussed in Section 2, public perceptions of GMFs are closely intertwined with nationalism, conspiracy thinking, and international geopolitics (Jiang & Fang, 2019; Xu et al., 2021). Consequently, for participants with high risk perception, their resistance and hesitation toward GMFs are likely rooted in deeper sociopolitical concerns and predispositions. This perspective helps explain why such predispositions anchor attitudes and, relative to the control, prevent information-focused personalization from yielding statistically greater change in a brief, single-session dialogue.

5.3. Moderating Roles of AI Literacy and Trust in Science

The moderation analysis differentiates two aspects of AI literacy: AI knowledge and AI experience. In the subgroup with negative risk perceptions, AI knowledge was associated with greater post-interaction attitude gains, whereas AI experience was associated with smaller gains. Substantively, this pattern points to a dual

pathway in AI-mediated persuasion. On the one hand, a basic conceptual understanding of AI may reduce uncertainty costs and facilitate message processing, thereby increasing receptivity to persuasive content. This interpretation aligns with prior research showing that greater AI-related knowledge is associated with higher levels of trust in AI systems (Shin, 2022; C. Wang et al., 2025; Yang & Wibowo, 2022), which may in turn function as a mediating factor enhancing persuasive effectiveness.

On the other hand, more extensive use of AI may raise evaluative thresholds or diminish novelty benefits, leading users to engage with AI-generated messages more critically and thereby decreasing marginal persuasion. Two mechanisms may account for this pattern. First, algorithm aversion theory suggests that exposure to algorithmic errors can rapidly erode confidence in algorithmic systems, prompting individuals to discount or resist algorithmic advice even when such systems outperform human judgment (Dietvorst et al., 2015). In the context of conversational AI, this means that greater prior experience may increase exposure to unsatisfactory or erroneous responses. In turn, this could reduce reliance on its advice and weaken persuasive effects. Second, while users increasingly rely on AI systems, greater familiarity is often accompanied by stricter expectations and heightened scrutiny (Lee et al., 2025). As frequent users are more likely to encounter AI imperfections, accumulated experience may heighten sensitivity to perceived failures, fairness concerns, or inconsistencies, resulting in more skeptical evaluations and, ultimately, weaker attitude change in AI-mediated dialogue (Jones-Jang & Park, 2023; Lucas et al., 2024).

Trust in science showed a different moderation profile. Across the full sample, higher trust corresponded to smaller marginal attitude shifts following interaction, and this attenuation was significant within the subgroup with more favorable prior views of GMFs, but not within the skeptical subgroup. A plausible explanation is a ceiling effect: Those who already trust scientific institutions and hold comparatively positive priors have limited headroom for additional movement (Sturgis et al., 2021). Another explanation lies in the potential misalignment between general trust in science and trust in specific scientific practices or topics. In other words, individuals may express confidence in science as a whole while maintaining reservations about particular domains. Prior research has described this group as the “critically interested” in science (Schäfer et al., 2018). In the context of our study, critically interested individuals may therefore help explain the moderation effects of trust in science—representing those who hold general trust in science yet remain critical of GMFs in particular. This observation is consistent with the aforementioned paradox of trust in China, where relatively high baseline trust in institutions coexists with persistent anxieties about GMFs (Cologna et al., 2025; Cui & Shoemaker, 2018; Huang et al., 2021).

5.4. Limitations and Future Work

This study has several limitations. Although the present experiment examined only short-term outcomes, it remains uncertain whether the observed effects can persist over time or translate into stable behavioral change. Future research should test whether these effects persist over time and translate into lasting behavioral change. In addition, individuals with lower educational attainment and non-urban populations are underrepresented in our sample, which may limit the generalizability of the findings to the broader population. The China-specific sociocultural context further constrains the external validity of the results beyond national boundaries. As Jin et al. (2022) note, public attitudes toward GMFs in China are characterized by enthusiasm for research and development alongside caution toward cultivation, substantial public opposition, and distrust of government regulation—factors that may shape attitudes and opinions in distinctive ways. Although we

demonstrate the effectiveness of personalization based on individual-level factors, persuasive processes are inevitably embedded within broader social and cultural contexts, which may condition how information is processed and received. Moreover, the present findings are derived from interactions conducted in Chinese with a Chinese-language conversational AI (DeepSeek). Linguistic framing and sociocultural norms embedded in Chinese-language communication may shape both the style of AI-generated responses (Lu, 2025; Ta et al., 2025) and users' reception of persuasive messages. These considerations highlight the need for cross-national and cross-cultural replications to assess the generalizability of our findings and to clarify the boundary conditions under which AI-based persuasion is most effective.

In addition, although we conducted fact-checking on a random sample of AI–user dialogues, we cannot guarantee that all AI-generated outputs were entirely error-free. Prior research suggests that such inaccuracies are a structural feature of LLMs (Shao, 2025; Y. Zhang et al., 2025). Future studies should therefore treat factual accuracy as an important boundary condition when evaluating the persuasive effects of AI-mediated communication. More importantly, in the context of personalized persuasion, factual inaccuracies may not be randomly distributed. Conditioning responses on user-specific attributes or beliefs may interact with model biases and introduce systematically skewed or selectively framed content. Recent studies (see Hackenburg et al., 2025; Lin et al., 2025) have raised concerns that AI-mediated persuasion may enhance persuasive effectiveness by sacrificing epistemic precision. While this possibility requires validation with larger samples and dedicated research designs, it was not part of the core research questions of the present study. We therefore view this potential accuracy–persuasion trade-off as an important direction for future research.

5.5. Conclusion

To sum up, the primary aim of this study was to investigate under which conditions AI-based persuasion performs better. We therefore manipulated the personalization level of persuasive efforts, and considered individual differences in AI literacy, trust in science, and risk perceptions. Using a four-arm between-subjects design, participants engaged in brief dialogues with a DeepSeek-based conversational AI.

Across all conditions, interaction with the conversational AI significantly improved both attitudes toward and willingness to consume GMFs, indicating that conversational AI-mediated dialogue itself possesses persuasive potential. Beyond this general effect, personalization enhanced persuasion only when demographic cues were combined with users' risk perceptions; neither demographics-only nor risk-perception-only tailoring produced reliable improvements relative to the control condition. This finding underscores that effective AI-based personalization requires the integration of both shallow audience characteristics and deeper, issue-proximal concerns. Persuasive effects were also heterogeneous across audience groups. Gains were concentrated among participants with comparatively favorable prior views of GMFs (i.e., lower perceived risk), whereas highly skeptical participants exhibited limited change. In addition, moderating effects diverged from conventional expectations: Greater use of AI and higher trust in science were associated with smaller marginal gains overall, while knowledge about AI facilitated persuasion only within the skeptical subgroup.

Together, these results specify when and how AI-based personalization is most effective and delineate boundary cases that call for alternative communication strategies. This study offers contextual empirical

insights into GMF communication in China, advancing theoretical understanding of how personalization shapes the persuasive potential of conversational AI. Empirically, the findings provide a basis for future work on leveraging conversational AI to foster public understanding of complex scientific issues across different sociopolitical contexts. Methodologically, the research design presented here is transferable to future explorations across a broader range of topics, enabling the testing of different models and the examination of AI-mediated persuasion in diverse cultural environments.

Acknowledgments

We are grateful to Xiaoyu Zhou for her valuable assistance and support throughout this study.

Conflict of Interests

The authors declare no conflict of interests.

Data Availability

The data include detailed chat records and therefore cannot be shared publicly to protect participant privacy. De-identified data may be made available by the corresponding author upon reasonable request and subject to ethical approval and data-sharing agreements.

An online appendix for this article is available online here: https://osf.io/8pz4y/overview?view_only=0f9c28c16d664ee9889abc41856eaa64

LLMs Disclosure

The authors did not use any LLMs for the conception of the study, experimental design, data collection, analysis, or the substantive writing of the manuscript. LLMs were used only for proofreading and minor grammatical edits.

References

- Alslaity, A., Chan, G., & Orji, R. (2023). A panoramic view of personalization based on individual differences in persuasive and behavior change interventions. *Frontiers in Artificial Intelligence*, 6, Article 1125191. <https://doi.org/10.3389/frai.2023.1125191>
- Bawa, A. S., & Anilakumar, K. R. (2013). Genetically modified foods: Safety, risks and public concerns—A review. *Journal of Food Science and Technology*, 50(6), 1035–1046. <https://doi.org/10.1007/s13197-012-0899-1>
- Brossard, D., & Nisbet, M. C. (2007). Deference to scientific authority among a low information public: Understanding U.S. opinion on agricultural biotechnology. *International Journal of Public Opinion Research*, 19(1), 24–52. <https://doi.org/10.1093/ijpor/edl003>
- Bucchi, M., & Trench, B. (2021). Rethinking science communication as the social conversation around science. *Journal of Science Communication*, 20(3), Article Y01. <https://doi.org/10.22323/2.20030401>
- Burright, J., & Al-khateeb, S. (2025). A comparative analysis of the ethics of gene editing: ChatGPT vs. Bard. *Computational and Mathematical Organization Theory*, 31(2), 195–206. <https://doi.org/10.1007/s10588-024-09392-z>
- Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS—Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies. *Computers in Human Behavior: Artificial Humans*, 1(2), Article 100014. <https://doi.org/10.1016/j.chbah.2023.100014>

- Chen, C., & Shu, K. (2024). Combating misinformation in the age of LLMs: Opportunities and challenges. *AI Magazine*, 45(3), 354–368. <https://doi.org/10.1002/aaai.12188>
- Chen, K., Shao, A., Burapachee, J., & Li, Y. (2024). Conversational AI and equity through assessing GPT-3's communication with diverse social groups on contentious topics. *Scientific Reports*, 14(1), Article 1561. <https://doi.org/10.1038/s41598-024-51969-w>
- Cologna, V., Mede, N. G., Berger, S., Besley, J., Brick, C., Joubert, M., Maibach, E. W., Mihelj, S., Oreskes, N., Schäfer, M. S., van der Linden, S., Abdul Aziz, N. I., Abdulsalam, S., Shamsi, N. A., Aczel, B., Adinugroho, I., Alabrese, E., Aldoh, A., Alfano, M., . . . Zwaan, R. A. (2025). Trust in scientists and their role in society across 68 countries. *Nature Human Behaviour*, 9, 713–730. <https://doi.org/10.1038/s41562-024-02090-5>
- Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, 385(6714), Article eadq1814. <https://doi.org/10.1126/science.adq1814>
- Cui, K., & Shoemaker, S. P. (2018). Public perception of genetically-modified (GM) food: A nationwide Chinese consumer study. *npj Science of Food*, 2, Article 10. <https://doi.org/10.1038/s41538-018-0018-4>
- Del Ponte, A., Li, L., Ang, L., Lim, N., & Seow, W. J. (2024). Evaluating SoJump.com as a tool for online behavioral research in China. *Journal of Behavioral and Experimental Finance*, 41, Article 100905. <https://doi.org/10.1016/j.jbef.2024.100905>
- Depounti, I., & Natale, S. (2025). Decoding artificial sociality: Technologies, dynamics, implications. *New Media & Society*, 27(10), 5457–5470. <https://doi.org/10.1177/14614448251359217>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>
- Du, Z., Xiao, Y., & Xu, J. (2022). How does information exposure affect public attitudes toward GMO in China? The mediating and moderating roles of conspiracy belief and knowledge. *Frontiers in Psychology*, 13, Article 955541. <https://doi.org/10.3389/fpsyg.2022.955541>
- Greussing, E., Guenther, L., Baram-Tsabari, A., Dabran-Zivan, S., Jonas, E., Klein-Avraham, I., Taddicken, M., Agergaard, T. E., Beets, B., Brossard, D., Chakraborty, A., Fage-Butler, A., Huang, C.-J., Kankaria, S., Lo, Y.-Y., Nielsen, K. H., Riedlinger, M., & Song, H. (2025). The perception and use of generative AI for science-related information search: Insights from a cross-national study. *Public Understanding of Science*, 34(5), 599–615. <https://doi.org/10.1177/09636625241308493>
- Hackenburg, K., & Margetts, H. (2024). Evaluating the persuasive influence of political microtargeting with large language models. *Proceedings of the National Academy of Sciences*, 121(24), Article e2403116121. <https://doi.org/10.1073/pnas.2403116121>
- Hackenburg, K., Tappin, B. M., Hewitt, L., Saunders, E., Black, S., Lin, H., Fist, C., Margetts, H., Rand, D. G., & Summerfield, C. (2025). The levers of political persuasion with conversational artificial intelligence. *Science*, 390(6777), Article eaea3884. <https://doi.org/10.1126/science.aea3884>
- Huang, Y.-H. C., Wang, X., Fong, I. W.-Y., & Wu, Q. (2021). Examining the role of trust in regulators in food safety risk assessment: A cross-regional analysis of three Chinese societies using an integrative framework. *Sage Open*, 11(4). <https://doi.org/10.1177/21582440211061579>
- Jiang, S., & Fang, W. (2019). Misinformation and disinformation in science: Examining the social diffusion of rumours about GMOs. *Cultures of Science*, 2(4), 327–340. <https://doi.org/10.1177/209660831900200407>
- Jin, Y., Schaub, S., Tosun, J., & Wesseler, J. (2022). Does China have a public debate on genetically modified organisms? A discourse network analysis of public debate on Weibo. *Public Understanding of Science*, 31(6), 732–750. <https://doi.org/10.1177/09636625211070150>

- Jones-Jang, S. M., & Park, Y. J. (2023). How do people react to AI failure? Automation bias, algorithmic aversion, and perceived controllability. *Journal of Computer-Mediated Communication*, 28(1), Article zmac029. <https://doi.org/10.1093/jcmc/zmac029>
- Kaptein, M., Markopoulos, P., de Ruyter, B., & Aarts, E. (2015). Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human-Computer Studies*, 77, 38–51. <https://doi.org/10.1016/j.ijhcs.2015.01.004>
- Karinshak, E., Liu, S. X., Park, J. S., & Hancock, J. T. (2023). Working with AI to persuade: Examining a large language model's ability to generate pro-vaccination messages. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1), Article 116. <https://doi.org/10.1145/3579592>
- Klimmt, C., & Spreen, N. (2025). Bettering biotech foods' foul flavor: An experiment on correcting citizens' misconceptions about genetic modification. *Science Communication*, 47(6), 767–799. <https://doi.org/10.1177/10755470241308532>
- Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Article 1121). Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713778>
- Lin, H., Czarnek, G., Lewis, B., White, J. P., Berinsky, A. J., Costello, T., Pennycook, G., & Rand, D. G. (2025). Persuading voters using human–artificial intelligence dialogues. *Nature*, 648(8093), 394–401. <https://doi.org/10.1038/s41586-025-09771-9>
- Liu, Y.-L., Yan, W., Hu, B., Li, Z., & Lai, Y. L. (2022). Effects of personalization and source expertise on users' health beliefs and usage intention toward health chatbots: Evidence from an online experiment. *Digital Health*, 8. <https://doi.org/10.1177/20552076221129718>
- Lu, H. (2025). Generative AI for vaccine misbelief correction: Insights from targeting extraversion and pseudoscientific beliefs. *Vaccine*, 54, Article 127018. <https://doi.org/10.1016/j.vaccine.2025.127018>
- Lucas, M., Zhang, Y., Bem-haja, P., & Vicente, P. N. (2024). The interplay between teachers' trust in artificial intelligence and digital competence. *Education and Information Technologies*, 29(17), 22991–23010. <https://doi.org/10.1007/s10639-024-12772-2>
- Marques, M. D., Kerr, J. R., Stukas, A. A., & McLennan, J. (2024). The effect of scientific conspiracy theories on scepticism towards biotechnologies. *Public Understanding of Science*, 34(5), 646–664. <https://doi.org/10.1177/09636625241300896>
- Master, Z., & Resnik, D. B. (2013). Hype and public trust in science. *Science and Engineering Ethics*, 19(2), 321–335. <https://doi.org/10.1007/s11948-011-9327-6>
- Matz, S. C., Teeny, J. D., Vaid, S. S., Peters, H., Harari, G. M., & Cerf, M. (2024). The potential of generative AI for personalized persuasion at scale. *Scientific Reports*, 14(1), Article 4692. <https://doi.org/10.1038/s41598-024-53755-0>
- Nelson, C. H. (2001). Risk perception, behavior, and consumer response to genetically modified organisms: Toward understanding American and European public reaction. *American Behavioral Scientist*, 44(8), 1371–1388. <https://doi.org/10.1177/00027640121956737>
- Ou, M., Ho, S. S., & Wijaya, S. A. (2025). Harnessing AI to address misinformation on cultivated meat: The impact of chatbot expertise and correction sidedness. *Science Communication*, 47(6), 859–896. <https://doi.org/10.1177/10755470251315097>
- Priest, S. H., Bonfadelli, H., & Rusanen, M. (2003). The 'trust gap' hypothesis: Predicting support for biotechnology across national cultures as a function of trust in actors. *Risk Analysis*, 23(4), 751–766. <https://doi.org/10.1111/1539-6924.00353>

- Ryan, C. D. (2014). Biotechnology communications, mythmaking and the media. In S. J. Smyth, P. W. B. Phillips, & D. Castle (Eds.), *Handbook on agriculture, biotechnology and development* (pp. 550–564). Edward Elgar Publishing. <https://doi.org/10.4337/9780857938350.00041>
- Salvi, F., Horta Ribeiro, M., Gallotti, R., & West, R. (2025). On the conversational persuasiveness of GPT-4. *Nature Human Behaviour*, 9, 1645–1653. <https://doi.org/10.1038/s41562-025-02194-6>
- Savaget, P., Chiarini, T., & Evans, S. (2018). Empowering political participation through artificial intelligence. *Science & Public Policy*, 46(3), 369–380. <https://doi.org/10.1093/scipol/scy064>
- Schäfer, M. S. (2023). The Notorious GPT: Science communication in the age of artificial intelligence. *Journal of Science Communication*, 22(2), Article Y02. <https://doi.org/10.22323/2.22020402>
- Schäfer, M. S., Füchslin, T., Metag, J., Kristiansen, S., & Rauchfleisch, A. (2018). The different audiences of science communication: A segmentation analysis of the Swiss population's perceptions of science and their information and media use patterns. *Public Understanding of Science*, 27(7), 836–856. <https://doi.org/10.1177/0963662517752886>
- Shao, A. (2025). New sources of inaccuracy? A conceptual framework for studying AI hallucinations. *Harvard Kennedy School Misinformation Review*, 6(4). <https://doi.org/10.37016/mr-2020-182>
- Shin, D. (2022). How do people judge the credibility of algorithmic sources? *AI & Society*, 37(1), 81–96. <https://doi.org/10.1007/s00146-021-01158-4>
- Sturgis, P., Brunton-Smith, I., & Jackson, J. (2021). Trust in science, social consensus and vaccine confidence. *Nature Human Behaviour*, 5(11), 1528–1534. <https://doi.org/10.1038/s41562-021-01115-7>
- Ta, N., Zeng, J., & Li, Z. (2025). Governance of discriminatory content in conversational AIs: A cross-platform and cross-cultural analysis. *Information, Communication & Society*. Advance online publication. <https://doi.org/10.1080/1369118X.2025.2537803>
- Teeny, J. D., & Matz, S. C. (2024). We need to understand “when” not “if” generative AI can enhance personalized persuasion. *Proceedings of the National Academy of Sciences*, 121(43), Article e2418005121. <https://doi.org/10.1073/pnas.2418005121>
- Tonui, W. K., Ogoyi, D., Thuo, C., Tareh, C., Alukhaba, C. L., Ouedraogo, A., Massouroudini, A., Regine, A. W. N., Ogbaki, A. G., Andae, G., Rotich, G., Kipkoech, A., & Omungo, R. (2024). Communicating the uncertainties associated with genetic biocontrol approaches: Insights from communicators, science journalists and scientists in Africa. *Journal of Science Communication*, 23(6), Article N02. <https://doi.org/10.22323/2.23060802>
- Wang, C., Boerman, S. C., Kroon, A. C., Möller, J., & de Vreese, C. H. (2025). The artificial intelligence divide: Who is the most vulnerable? *New Media & Society*, 27(7), 3867–3889. <https://doi.org/10.1177/14614448241232345>
- Wang, G., Wang, L., & Shen, J. (2021). Food to politics: Representations of genetically modified organisms in cartoons on the internet in China. *Public Understanding of Science*, 30(3), 274–284. <https://doi.org/10.1177/0963662520983564>
- Wu, J., Wang, Z., & Qin, Y. (2025). Performance of DeepSeek-R1 and ChatGPT-4o on the Chinese National Medical Licensing Examination: A comparative study. *Journal of Medical Systems*, 49(1), Article 74. <https://doi.org/10.1007/s10916-025-02213-z>
- Xia, D., Song, M., & Zhu, T. (2024). A comparison of the persuasiveness of human and ChatGPT generated pro-vaccine messages for HPV. *Frontiers in Public Health*, 12, Article 1515871. <https://doi.org/10.3389/fpubh.2024.1515871>
- Xu, Q., Song, Y., Yu, N., & Chen, S. (2021). Are you passing along something true or false? Dissemination of social media messages about genetically modified organisms. *Public Understanding of Science*, 30(3), 285–301. <https://doi.org/10.1177/0963662520966745>

- Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, 8(10), 1839–1850. <https://doi.org/10.1038/s41562-024-02004-5>
- Yang, R., & Wibowo, S. (2022). User trust in artificial intelligence: A comprehensive conceptual framework. *Electronic Markets*, 32(4), 2053–2077. <https://doi.org/10.1007/s12525-022-00592-6>
- Yuan, S., Ma, W., & Besley, J. C. (2019). Should scientists talk about GMOs nicely? Exploring the effects of communication styles, source expertise, and preexisting attitude. *Science Communication*, 41(3), 267–290. <https://doi.org/10.1177/1075547019837623>
- Zhang, W., Xue, J., Folmer, H., & Hussain, K. (2019). Perceived risk of genetically modified foods among residents in Xi'an, China: A structural equation modeling approach. *International Journal of Environmental Research and Public Health*, 16(4), Article 574. <https://doi.org/10.3390/ijerph16040574>
- Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Xu, C., Chen, Y., Wang, L., Luu, A. T., Bi, W., Shi, F., & Shi, S. (2025). *Siren's song in the AI ocean: A survey on hallucination in large language models*. arXiv. <https://doi.org/10.48550/arXiv.2309.01219>
- Zhou, T., & Li, S. (2024). Understanding user switch of information seeking: From search engines to generative AI. *Journal of Librarianship and Information Science*. Advance online publication. <https://doi.org/10.1177/09610006241244800>

About the Authors



Qi Xi is a PhD candidate at Beijing Normal University. His research interests include AI persuasion and computational communication.



Jing Zeng is an assistant professor of computational communication and social science at the Department of communication and media research (IKMZ), University of Zurich.



Zhanghao Li is an assistant professor at the School of Journalism and Communication, Guangzhou University. His research interests include algorithm auditing, AI persuasion, and computational communication.



Mike S. Schäfer is a full professor of science communication at the Department of Communication and Media Research (IKMZ), University of Zurich.