Simulated Misinformation Susceptibility (SMISTS): Enhancing Misinformation Research with Large Language Model Simulations

Weicheng Ma[1], **Chunyuan Deng**[2], **Aram Moosavi**[3], **Lili Wang**[3], **Soroush Vosoughi**[3], and **Diyi Yang**[4]

[1,3]Department of Computer Science, Dartmouth College
[2]College of Computing, Georgia Institute of Technology
[4]Computer Science Department, Stanford University
[1]weicheng.ma.gr@dartmouth.edu
[4]diyiy@stanford.edu

Abstract

Psychological inoculation, a strategy to build resistance against persuasive misinformation, has been shown to reduce its spread and adverse effects. Although these inoculations are effective, the design and optimization of them typically require substantial financial and human resources. To address these challenges, this work introduces Simulated Misinformation Susceptibility Test (SMIST), leveraging Large Language Models (LLMs) to simulate participant responses in misinformation studies. SMIST employs a life experience-driven simulation methodology, which accounts for various aspects of participants' backgrounds, to mitigate common issues of caricatures and stereotypes in LLM simulations and enhance response diversity. Our extensive experimentation demonstrates that SMIST, utilizing GPT-4 as the backend model, yields results that align closely with those obtained from human-subject studies in misinformation susceptibility. This alignment suggests that LLMs can effectively serve as proxies in evaluating the impact of psychological inoculations. Further, SMIST can be applied to emerging and anticipated misinformation scenarios without harming human participants, thereby expanding the scope of misinformation research.

1 Introduction

Misinformation poses significant, far-reaching harm, such as fueling vaccination hesitancy and exacerbating long-term mental health issues, notably during the COVID-19 pandemic (Do Nascimento et al., 2022). Differing from other social issues, early intervention is crucial in combating misinformation, as post-hoc corrections often fail to mitigate its detrimental effects (Nyhan et al., 2014). Psychological inoculations, designed to weaken the belief in and spread of misinformation, are effective but require extensive Misinformation Susceptibility Tests (MISTs) during development (Roozenbeek et al., 2022; Traberg et al., 2022; Basol et al., 2020). MISTs are designed to evaluate individual susceptibility to misinformation, shedding light on the factors that contribute to vulnerability. We follow prior studies in our definition of susceptibility (Basol et al., 2020; Roozenbeek et al., 2020), which use analyses of variance (ANOVA) and regression analyses on participant responses to MISTs to measure susceptibility to misinformation as a predictor of belief in misinformation (we provide more detail in later sections). These tests usually involve subjecting participants to diverse misinformation forms to gauge their propensity to believe and disseminate such information. However, MISTs are hindered by significant labor and time demands, along with challenges in reproducibility and lim-

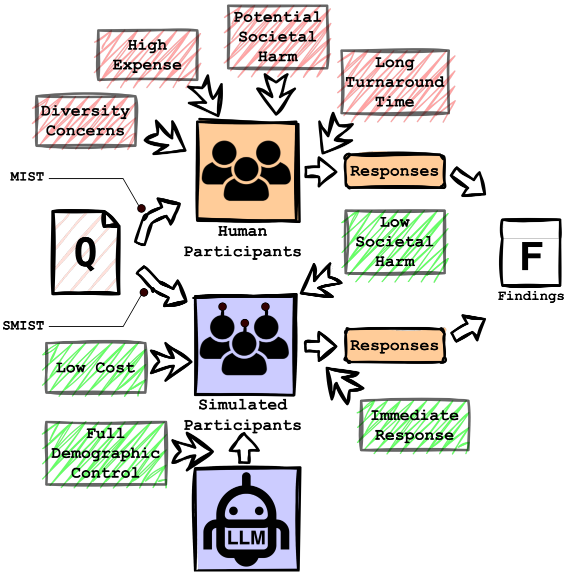

Figure 1: Comparative illustration of the processes in human-subject MIST (top) and SMIST (bottom). Red and green boxes demonstrate cons of MIST and pros of SMIST, respectively. Q indicates the questionnaire. In SMIST, while all simulated participants within a group share specified demographic characteristics, they exhibit diversity in other uncontrolled aspects.

Findings of the Association for Computational Linguistics: ACL 2024, pages 2774–2788
August 11-16, 2024 ©2024 Association for Computational Linguistics

ited generalizability across various misinformation types. Additionally, they pose an ethical dilemma by potentially exposing the public to harmful misinformation before its widespread dissemination. Leveraging recent advancements in large language models (LLMs) and their increasing application in computational social science (Ziems et al., 2023; Argyle et al., 2023), we propose the use of simulated MISTs (SMISTs), utilizing LLMs in place of human subjects to address these challenges. Although simulating LLMs as agents is a new and emerging topic, recent work has started to explore this concept in various scientific domains (Park et al., 2023; Shanahan et al., 2023; Wang et al., 2023; Xi et al., 2023). These inquiries motivate our use of LLMs to study misinformation susceptibility. Figure 1 illustrates the SMIST methodology compared to traditional human-subject MISTs.

SMIST involves two main stages: participant profile generation and questionnaire response, both predominantly managed by LLMs. During the profile generation phase, we utilize an LLM to fabricate participant life experiences based on a specific set of demographic features. To mitigate the tendency of LLMs to produce stereotypical or caricatured profiles, we limit the input to one demographic feature per profile and encourage the model to improvise on other demographic aspects (see Appendix A for an example prompt and a generated life story). This approach has proven effective in generating a rich diversity in the uncontrolled dimensions of the profiles while ensuring that the controlled demographic features are represented accurately and without stereotyping, as confirmed by our manual reviews and validations. For the questionnaire response stage, we align with the methodology described in Argyle et al. (2023). Here, the LLM is instructed to answer misinformation questionnaires, adopting the perspectives of the participants with the created life experiences (see Figure 4 for an example prompt).

Our initial application of the SMIST focused on COVID-19 misinformation. In these tests, SMIST demonstrated robustness and reproducibility, akin to human participants in MISTs, with consistent results across repeated queries and various sample sizes. We employed statistical methods like ANOVA and ordinary least squares (OLS) regression to explore the relationship between demographic features of simulated participants and their susceptibility to misinformation. The outcomes corroborate findings from prior COVID-19 misinformation studies (e.g., Roozenbeek et al. (2020)) and our human-subject experiments. Further, we expanded our experiments to encompass different misinformation forms (satire, conspiracy theories), topics (Russo-Ukraine war, 2020 US election, GMO, climate change), and time frames (recent vs. past misinformation). In each of these contexts, SMIST's outputs remained in agreement with both our human-based studies and existing literature (e.g., Uscinski et al. (2020)). This consistency reinforces SMIST's capability to replicate human-subject MIST results, its capability for controlled demographic representation, and its generalizability. Significantly, SMIST effectively circumvents the ethical concern of exposing the public to misinformation prematurely, a notable issue in traditional MISTs. This positions SMIST as a valuable tool for developing rapid, effective psychological inoculations against misinformation. This paper's contributions are twofold:

Development of SMIST: We introduce SMIST, a novel approach aiming to enhance the efficiency, reproducibility, and generalizability of human-subject MISTs, facilitating the design of psychological inoculations against misinformation.

Life Experience-Based Simulation with LLMs: We refine and extend the application of life experience-based simulations using LLMs to create more nuanced participant profiles. This method significantly reduces the risk of caricatures and stereotypes often present in LLM simulations, presenting broader implications for simulation-intensive research domains.

2 Misinformation Data Preparation

In our research, misinformation was sourced from three distinct reputable channels: (1) news articles from The New York Times[1] and USA Today[2] reporting on recent misinformation about various topics; (2) Wikipedia entries detailing instances of misinformation, exemplified by the entry "Disinformation in the Russian invasion of Ukraine"[3]; and (3) organizations dedicated to refuting misinformation, such as the Alliance for Science[4]. To ensure authenticity, we verified that each misinformation item was genuinely circulating online rather than

[1]https://www.nytimes.com/

[2]https://www.usatoday.com/

[3]https://en.wikipedia.org/wiki/Disinformation_in_the_Russian_invasion_of_Ukraine

[4]https://allianceforscience.org/

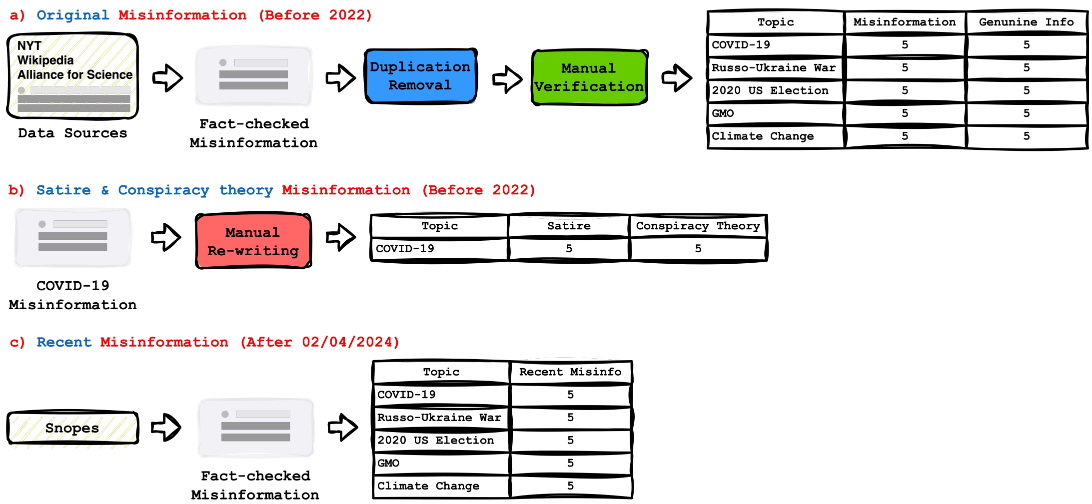

Figure 2: The construction process of (a) an original fact-checked misinformation & genuine information dataset, (b) a COVID-19-related satire & conspiracy theory dataset, and (c) a recent misinformation dataset containing misinformation fact-checked on Snopes after 02/04/2024. The number of (mis)information pieces under each category is also displayed.

being fabricated by a single source. Our collection spanned five diverse topics: COVID-19, the Russo-Ukraine war, the 2020 US election, GMOs, and climate change. Each misinformation piece was accompanied with a corresponding explanation and correction, which were used to create a genuine information corpus for the topics. Additionally, misinformation related to COVID-19 was creatively transformed into satire and conspiracy theory versions by a domain expert in our group, who is a native English speaker. To be specific, we exaggerated the absurdity of claims by identifying and amplifying their core elements to create satire-styled misinformation. For conspiracy-based misinformation, we used social media data to link events or individuals with specific misinformation, then modified these narratives to suggest associations with the identified subjects. The transformations applied in our study are limited to COVID-19-related misinformation, owing to the scarcity of research on people's susceptibility to misinformation characterized by satire and conspiracy theories in other subjects. These adaptations underwent a quality assurance review. Within our experimental framework, we selected five instances of misinformation for each topic, including their genuine, satirical, and conspiratorial versions (if available).

Moreover, we incorporated five recent fact-checked misinformation pieces from Snopes[5] per topic. All these recent misinformation pieces have been fact-checked on Snopes after 02/05/2024, less than a week before our human-subject MIST questionnaires were distributed. This inclusion of recent misinformation aimed to test SMIST's robustness against unfamiliar misinformation (i.e., recent and not widely spread misinformation). Moreover, using recent misinformation allows us to validate our human-subject MIST, given the potential refusal of both humans and LLMs to trust or share widely recognized misinformation.

Figure 2 illustrates the construction and sizes of datasets utilized in our study. We refer to the three datasets as (a) original, (b) satire and conspiracy theory, and (c) recent misinformation datasets.

3 Methodology of SMIST

This section discusses the methodology of SMIST, focusing on two key aspects. First, we describe the creation of participant profiles using an LLM, ensuring a balanced representation of demographic features (Section 3.2). Second, we explain how these profiles guide an LLM in responding to MIST questionnaires, encapsulating the simulated participants' perspectives (Section 3.1). The comprehensive SMIST process is visually summarized in Figure 3. We use OpenAI's GPT-3.5 for generating participant life experiences and GPT-4 for answering the questionnaires. Our choice of GPT-4 for questionnaire answering stems from preliminary analyses that evaluated several LLMs including GPT-3.5, GPT-4, and Google BARD. GPT-4 emerged as the superior model in simulating hu-

[5] https://www.snopes.com/fact-check/

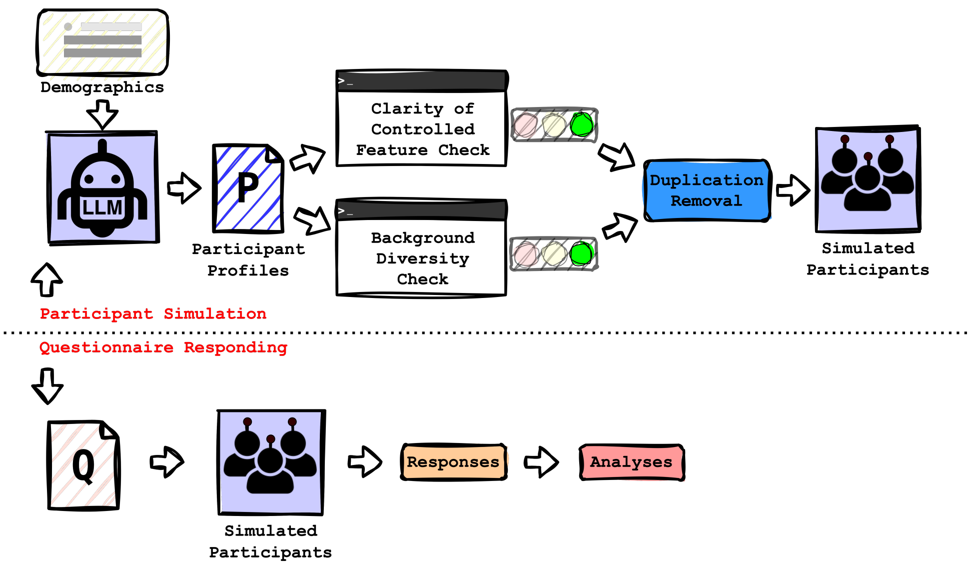

Figure 3: The participant simulation and questionnaire response stages involved in SMIST.

```
life story: [Aisha Khan: Hailing from Atlanta,
Georgia, Aisha is a 38-year-old African-
American woman with a strong entrepreneurial
spirit. After obtaining her MBA, she founded a
successful tech startup focused on empowering
underprivileged communities through education.
Aisha is a practicing Muslim and actively
participates in interfaith dialogue. She
believes in the importance of a well-informed
citizenry and trusts the media to provide
accurate information. Aisha supports government
initiatives that promote diversity, equality,
and economic empowerment.]                          } Life story

Suppose you are Aisha with your life experience
described above, rate your willingness to share
the information [Dominion, a company that makes
electronic voting machine, rigged the 2020
election against Donald Trump].                     } Questionnaire

Answer in 7-point Likert scale and explain your
rating.
```

Figure 4: An example SMIST questionnaire.

man responses to MISTs. The extensive time and resources required for experimenting with LLMs led us to focus our comprehensive set of experiments solely on GPT-4.

3.1 Questionnaire Design

In MISTs, two prevalent types of questionnaires are utilized: (1) information accuracy tests, where participants assess the accuracy of a given piece of (mis)information, and (2) willingness-to-share tests, inquiring about the participants' likelihood of sharing the information (Arin et al., 2023). While belief in information accuracy is intuitively more directly associated with misinformation susceptibility, both types of questionnaires have been examined in prior studies (Chua and Banerjee, 2018). As such, our study employs both questionnaire types to ensure comprehensive analysis. Specifically, we instruct the GPT-4 model to simulate a participant's perspective, evaluating the accuracy or sharing propensity of the (mis)information. Due to the lack of a universally accepted scoring scale, we adopt the 7-point Likert scale as per Roozenbeek et al. (2020) for a consistent comparison with their results. An example SMIST questionnaire is provided in Figure 4. The questionnaire includes a "life story" section to maintain diversity among participants with identical controlled demographic characteristics, as detailed in Section 3.2. The second component of the questionnaire comprises the actual questionnaire items, querying perceptions of information accuracy or the likelihood of sharing the information.

3.2 Participant Simulation

A known challenge with LLMs simulating human behaviors is their propensity to overemphasize characteristics in the prompts, leading to homogeneous or stereotypical responses (Cheng et al., 2023). To address this and promote response diversity in SMIST, we integrate participants' life experiences into the simulation prompts, as suggested by Ma et al. (2023). Our methodology involves providing only one demographic feature per life story to maximize diversity. The demographic attributes in our study encompass age (young adult, middle-aged, elderly), gender (male, female), education level (non-educated, secondary school diploma, bachelor's degree), political leaning (left wing, right wing), and trust in scientists, government, and journalism (trust, distrust). These demographic features are selected based on those identified by Roozenbeek et al. (2020), while features uniquely related to COVID-19 are omitted to broaden our research scope to other topics. It is important to note that this demographic feature set is not exhaustive. SMIST can be adapted to include various demographic and psychographic groups.

We conducted thorough manual checks to ensure that (1) the specified demographic features were accurately reflected in the life stories and (2) a high level of diversity was maintained in the other uncontrolled demographic features. This approach effectively mitigates stereotypical associations, such as linking specific genders to certain occupations (Kotek et al., 2023). Further, we annotated the uncontrolled demographic features of each life story to enrich our analysis, particularly examining the correlation between these features and the controlled demographic attribute. The results demonstrate significant profile diversity, with Pearson correlation coefficients ranging from -0.08 to 0.44. This diversity is crucial in preventing the LLM from disproportionately focusing on the controlled demographic feature during the questionnaire response phase.

Our participant information dataset comprises 160 life stories, with 10 for each of the 16 controlled demographic features. Appendix A includes an example life story with the "age group" set as "elderly," along with its manually annotated demographic features.

In structuring LLMs to mimic human behaviors, multiple methods are available, such as employing narrative descriptions of demographic characteristics or adopting an interview-style dialogue. Our preliminary tests explored three such strategies as outlined by Argyle et al. (2023), i.e., using narrative self-introductions, keyword-based self-introductions, or interview text to instruct the LLM participant characteristics. However, these methods frequently resulted in the model either refusing to respond or providing extreme or ambiguous scores, a consequence likely stemming from hallucinatory responses, which are unsuitable for the requirements of SMIST.

Specifically, our preliminary analysis involved querying the GPT-4 model with various participant simulation techniques, using a selection of three misinformation and three genuine information pieces from our dataset. These included information accuracy and willingness-to-share tests, employing a subset of five randomly selected participants from our pool. In the willingness-to-share assessments, our life story-based approach yielded no refusals to respond, with only a single instance (out of 30 tests) of slightly ambiguous scoring (i.e., a score of 5-6 on the 7-point Likert scale). In the information accuracy tests, ambiguous responses occurred in 2 out of 30 cases, with no refusals to answer noted. By contrast, the three previously studied participant simulation methods (Argyle et al., 2023) led to considerable refusal rates in both test types, ranging from 8 to 17 out of 30 cases in willingness-to-share tests and 6 to 15 in information accuracy tests. Additionally, ambiguous scores were observed in 13 to 14 cases for willingness-to-share tests and 1 to 14 cases for information accuracy tests, with some responses deviating textually from the requested numerical format. Overall, the application of the three approaches introduced by Argyle et al. (2023) led to refusal-to-answer rates of around 20% - 50% and ambiguous response rates above 50%. Due to the substantial rate of non-responsiveness or unclear answers, conducting further statistical analyses is impractical and could be misleading. Consequently, we disregarded these prompting approaches based on self-introduction or interview text and opted for the life-story approach to inform the model about the participants it simulates.

4 SMIST Results and Analyses

This section examines a few key research questions related to SMIST, around LLMs' ability to interpret participant profiles and respond to questionnaires, and how different demographic attributes affect misinformation susceptibility. Unless otherwise specified, the experiments are conducted using the original dataset as introduced in Section 2.

4.1 Robustness of SMIST Results

Prior to delving into the results, we first establish the robustness of SMIST against repeated queries. This step is crucial to ensure that the model's responses are not products of random hallucination. We conducted SMIST thrice for each set of (mis)information pieces and participants. This process yielded 16,000 scores per topic for each questionnaire type (10 participants $\times$ 5 (mis)information pieces $\times$ 2 categories: genuine information and misinformation $\times$ 2 questionnaire types $\times$ 5 topics $\times$ 16 controlled demographics). To assess consistency, we computed the Pearson and Spearman correlation coefficients for the re-

Controlled Demographics	COVID-19		Russo-Ukraine War		2020 US Election		GMO		Climate Change	
	Genuine Info	Misinfo	Genuine Info	Misinfo	Genuine Info	Misinfo	Genuine Info	Misinfo	Genuine Info	Misinfo
Information Accuracy Test Results										
Age	**4.23**	1.67	0.54	0.99	0.79	1.91	0.94	2.40	0.94	2.40
Gender	0.05	0.17	1.73	0.23	**9.49**	0.13	0.21	0.11	0.21	0.11
Education Level	2.95	2.74	**11.35**	**20.22**	1.33	**12.27**	0.55	**39.60**	0.55	**39.60**
Political Ideology	2.32	2.67	3.87	**7.95**	0.98	0.07	2.76	**6.11**	2.76	**6.11**
Trust in Scientists	0.48	0.14	0.54	0.01	**5.61**	0.01	0.35	0.80	0.35	0.80
Trust in Government	0.01	0.04	1.56	3.70	**5.09**	4.56	**5.65**	**10.95**	**5.65**	**10.95**
Trust in Journalism	0.35	0.18	5.08	2.13	2.21	0.00	0.36	0.28	0.36	0.28
Willingness-to-Share Test Results										
Age	**6.11**	1.66	**4.86**	2.01	2.63	**10.06**	**3.36**	**5.84**	**3.36**	**5.84**
Gender	0.02	0.18	0.87	3.76	**15.09**	1.01	**6.96**	0.40	**6.96**	0.40
Education Level	**4.40**	**10.65**	3.21	1.93	**28.35**	**11.92**	**5.64**	**16.17**	**5.64**	**16.17**
Political Ideology	1.08	0.81	0.00	0.03	0.14	0.94	**6.29**	3.79	**6.29**	3.79
Trust in Scientists	3.01	**23.24**	1.59	**43.58**	1.94	**83.07**	0.80	**48.19**	0.80	**48.19**
Trust in Government	1.07	0.12	**10.64**	**7.80**	**6.76**	**21.91**	1.39	0.29	1.39	0.29
Trust in Journalism	0.45	6.34	0.00	0.63	0.06	**6.70**	1.05	3.60	1.05	3.60

Table 1: ANOVA Results: f-values for Controlled Demographic Features as Independent Variables (Significant values are bolded).

Topic	Approach	Questionnaire	Intercept	Age	Gender (female)	Education Level	Political Ideology (conservative)	Trust in Scientists	Trust in Government	Trust in Journalism
COVID-19	SMIST	accuracy	2.04	-0.30	-0.18	-0.08	0.03	-0.10	-0.01	0.06
		share	1.22	-0.72	-0.44	-0.32	0.02	0.01	0.08	0.01
	Roozenbeek et al. (2020)	accuracy	2.30	-0.21	-0.16	0.04	0.09	-0.24	-0.01	0.11
	Pickles et al. (2021)	accuracy	NA	-0.02	-0.38	NA	NA	NA	-0.14	NA
Russo-Ukraine War	SMIST	accuracy	2.14	0.30	0.18	-0.08	0.03	-0.01	0.01	-0.06
		share	1.52	0.78	0.17	-1.12	0.06	0.05	0.18	0.04
2020 US Election	SMIST	accuracy	3.09	-0.07	-0.09	-0.23	0.06	0.00	-0.08	0.03
		share	3.06	0.64	-0.12	-0.70	0.05	0.05	-0.13	0.19
GMO	SMIST	accuracy	3.68	0.60	1.00	-0.25	0.05	0.03	0.03	-0.17
		share	2.80	-0.02	0.19	-0.76	-0.03	0.07	0.09	-0.35
Climate Change	SMIST	accuracy	3.98	0.55	-0.08	-0.12	-0.04	0.00	-0.04	0.05
		share	2.73	0.04	-0.27	-0.26	-0.06	0.02	-0.03	0.05

Table 2: The OLS regression coefficients from SMIST results on misinformation, alongside comparisons with two preceding studies that conducted similar regression analyses. 95% confidence intervals are shown in brackets.

peated trials, controlling for topic and questionnaire type. The coefficients consistently registered above 0.89 (Pearson) and 0.92 (Spearman) across different runs, with p-values below 0.05, often reaching below 0.01. These high correlation values indicate a significant consistency in the scores across repeated runs, strongly suggesting that the model's responses are not random but are reliably based on the simulated participant profiles. Additionally, we conducted preliminary experiments to test the sensitivity of SMISTs to minor prompt variations, e.g., comparing model responses when instructing it to "respond using a 7-point Likert scale" or to "specify your answer with a score ranging from 1 to 7". The SMIST results are shown to be stable given such expression-level differences in prompts. These results support the robustness of SMISTs and the credibility of using the model responses in research contexts.

4.2 Discussion of SMIST Results

In our analysis presented in Appendix B, we observe that *the mean scores vary significantly across demographic groups, while the in-group standard deviations remain low (between 0.10 and 0.60).* Notably, *in some instances, belief in misinformation approaches or exceeds that in genuine information.* These findings indicate that SMIST results authentically reflect participant characteristics, both controlled and uncontrolled, rather than being uniform and solely dependent on the LLM's knowledge. Otherwise, we would expect to see extreme and definitive scores (such as 1 for misinformation and 7 for genuine information, with very low standard deviations) for familiar (mis)information, and highly variable scores for unfamiliar (mis)information.

Moreover, SMIST demonstrates sensitivity to questionnaire design, evident in occasional discrepancies between information accuracy and willingness-to-share test outcomes. For instance, participants with a "distrust in scientists" profile show higher susceptibility to misinformation about the Russo-Ukraine war (mean score of 5.30), yet exhibit reluctance to share this misinformation (mean sharing score of 2.20). This aligns with existing literature (Arin et al., 2023) showing that *individuals may not always share misinformation they believe in.* Such inconsistencies in SMIST results underscore the importance of using both types of questionnaires in MISTs and affirm the adaptability of SMIST to various questionnaire formats.

4.3 Analyses of Variance

In our study, we employ one-way ANOVA, using the controlled demographic features as independent variables, to *assess the variability of scores across and within demographic groups* (e.g., among participants all classified as young adults). As indicated in Table 1, the f-values in ANOVA exhibit a broad range, from as low as 0.02 to as high as 83.07. As results in Section 4.2 suggest that SMIST scores are diverse but not random, higher f-values in ANOVA can be interpreted as indicative of certain demographic features being more influential in determining participants' trust in genuine information or susceptibility to misinformation, showing that SMIST effectively identifies the relative significance of different demographic features in predicting misinformation susceptibility. Our analysis aligns with similar research in prior studies (Roozenbeek et al., 2020), with our results showing comparable magnitudes to their findings on human-subject MIST responses, which bolsters the credibility of SMIST. Moreover, the presence of low f-values indicates *notable in-group score variances, even among participants sharing a controlled demographic trait*. This observation implies that our life experience-based participant simulation approach effectively mitigates the LLM's tendency to overemphasize prominent participant characteristics. Instead, the model appears to give due weight to a range of uncontrolled demographic features, thereby aligning more closely with the behavior of real participants in human-subject experiments.

4.4 OLS Regression Analyses

We employ OLS regression to further *explore the relationship between demographic features and participants' susceptibility to misinformation*. We considered seven demographic attributes—age, gender, education level, political ideology, trust in scientists, government, and journalism—as independent variables, analyzing their influence on the responses in SMISTs. The regression coefficients for these variables are detailed in Table 2. Our analysis revealed several key insights about SMIST:

Variability in Responses Across Questionnaire Types: *There are notable differences in responses between the information accuracy and willingness-to-share questionnaires.* E.g., an increase in age correlates with a decreased willingness to share and heightened belief in GMO-related misinformation, while the opposite trend is observed in misinformation related to the 2020 US election. This reflects the complexity of human behavior in MISTs, as documented in previous studies (Arin et al., 2023), suggesting that SMIST responses mirror the diverse perspectives of the simulated participants.

Alignment with Previous Studies: *Our OLS regression results largely align with findings from prior research.* Consistent with Roozenbeek et al. (2020), factors like "age" and "female gender" show a strong negative correlation with misinformation susceptibility, whereas "conservative political ideology" and "trust in journalism" are associated with increased susceptibility. While Pickles et al. (2021) observed a similar negative association between "female gender" and susceptibility, they found "age" to have negligible effects. *It is important to recognize that in the realm of misinformation research, consensus on the influence of various demographic features is not always established. This variability in the literature means that discrepancies between our SMIST findings and previous studies are not necessarily detrimental to SMIST's validity.* For instance, while we observed a negative correlation between "education level" and misinformation susceptibility, conflicting findings exist in the literature. Some studies report no significant association (Lunz Trujillo et al., 2021; Lyons et al., 2019), while others have found a negative correlation (De Coninck et al., 2021; Han et al., 2020), echoing our results. This context underscores the complexity and ongoing nature of research in this field. For the other four topics we examined, the relationships between demographic features and misinformation susceptibility derived from SMISTs also resonate with existing literature. For example, having a college degree is negatively associated with susceptibility to 2020 US election misinformation, in line with (Calvillo et al., 2021), while being politically right-winged correlates with higher susceptibility. Additionally, in line with the findings of Lai et al. (2022), our results corroborate the view that "education level" significantly predicts a lower susceptibility to climate change misinformation. Similarly, "age" emerges as a notable predictor of higher misinformation susceptibility in the accuracy tests, though this trend is not mirrored in the willingness-to-share tests. These parallels reinforce the credibility and relevance of our research methodology and findings in the broader context of misinformation studies.

Controlled Demographics	ANOVA f-value		OLS Coefficients	
	Satire	Conspiracy	Satire	Conspiracy
Information Accuracy Test Results				
(Intercept)			2.63	2.39
Age	0.10	2.85	-0.08	-0.01
Gender	1.09	1.04	-0.03	-0.04
Education Level	1.04	0.81	0.00	-0.10
Political Ideology	1.06	3.70	0.10	0.02
Trust in Scientists	0.79	14.68	-0.06	0.00
Trust in Government	0.80	0.98	0.41	0.03
Trust in Journalism	**28.85**	0.32	-0.05	0.05
Willingness-to-Share Test Results				
(Intercept)			1.19	1.52
Age	**2.87**	**4.59**	-0.31	-0.11
Gender	0.89	0.54	-0.05	-0.06
Education Level	0.96	2.33	0.08	-0.05
Political Ideology	0.54	4.09	0.06	0.05
Trust in Scientists	4.00	**9.56**	-0.18	-0.07
Trust in Government	0.03	2.04	0.06	0.13
Trust in Journalism	0.58	0.47	0.13	0.04

Table 3: SMIST Results on Satire and Conspiracy Theory Misinformation: ANOVA f-values and OLS Regression Coefficients (Significant f-values Bolded, 95% Confidence Intervals in Brackets).

Addressing Discrepancies in Literature: The literature exhibits varying opinions on the influence of certain demographic features on misinformation susceptibility. For instance, Calvillo et al. (2020) contradict the findings of Calvillo et al. (2021) regarding the impact of political orientation on susceptibility to election-related misinformation. Apart from these contested areas, SMIST's results generally align with existing literature, indicating its effectiveness in eliciting human-like responses. We speculate that discrepancies in previous research findings may arise from variations in the misinformation content used, the demographic distribution of participants (including uncontrolled or unstudied characteristics), and participants' knowledge about the misinformation. SMIST offers a practical, robust, and efficient approach for replicating and comparing studies under controlled conditions. By managing all variables, including questionnaires and participant profiles, SMIST facilitates fair comparisons and may help resolve existing inconsistencies in the field's findings.

4.5 SMIST's Generalization to Diverse Misinformation Forms

To assess SMIST's generalizability across various misinformation forms, *we expanded our experimentation to include satire- and conspiracy theory-styled misinformation*, focusing on COVID-19 to maintain consistency with earlier studies. This set of experiments are conducted over 10 pieces of misinformation (5 satire and 5 conspiracy theory) with 160 participants (16 controlled demogrpahics with 10 participants per group) and 2 questionnaire types, yielding a total of 3,200 scores. *The observed standard deviations for both questionnaire types remained low (typically under 0.5 and always below 1.01),* indicating SMIST's robustness irrespective of the misinformation format.

SMIST responses to these alternate forms of misinformation show notable deviation from those using original misinformation pieces, aligning with Gemenis (2021). We observed Pearson correlation coefficients ranging from -0.12 to 0.42 for each participant, indicating weak negative to moderate positive correlations. Such variations are evident in both ANOVA and OLS regression results (Table 3), underscoring SMIST's sensitivity to changes in misinformation forms.

Our analysis highlights key predictors of susceptibility to COVID-19 conspiracy theories, in line with Uscinski et al. (2020). These include a slight positive correlation between conservative political ideology and conspiracy beliefs, a minor negative impact of education level, and a more pronounced negative association between age and conspiracy beliefs. Contrasting with Uscinski et al. (2020)'s findings, our study identifies a slight negative correlation between gender and belief in COVID-19 conspiracy theories, more closely aligning with Gemenis (2021).

The results from SMIST are intuitive and align with expectations, even in areas lacking extensive literature for comparison. For example, a strong trust in scientists typically correlates with better discernment of misinformation. Additionally, *discrepancies between trusting and willingness to share misinformation are more pronounced in satirical and conspiratorial forms than in original misinformation, likely due to their more engaging and shareable nature.*

These findings affirm SMIST's strong generalizability across different misinformation forms, such as satire and conspiracy theory. Given the nascent state of research in this area, SMIST emerges as a valuable tool for enhancing the efficiency and scope of such studies.

4.6 Adapting SMIST to Recent Misinformation

In contrast to other research, the credibility of MIST results can be compromised if participants are already well-informed about the facts related to the misinformation used in the tests. To address this, we extend SMIST and human-subject MISTs to recently fact-checked misinformation as intro-

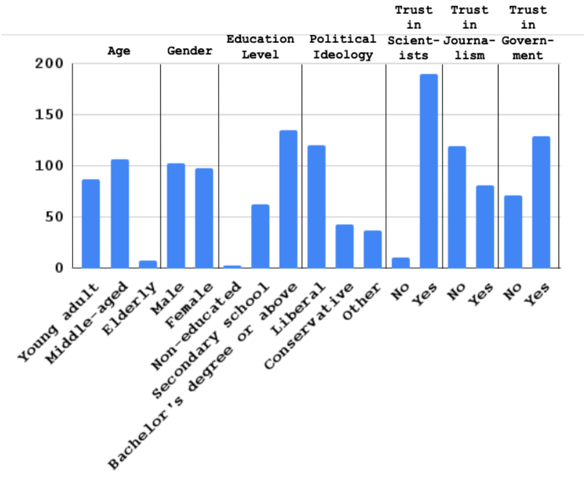

Figure 5: The distribution of demographic features in the participants of our human-subject MIST.

duced in Section 2, demonstrating SMIST's effectiveness with misinformation unfamiliar to its backend model. Specifically, we experimented with 5 pieces of recently fact-checked misinformation per topic for 5 topics, 160 participants (16 controlled demographics with 10 simulated participants per group), and 2 types of questionnaires, resulting in a total of 8,000 scores. We distributed 200 questionnaires via CloudResearch.com. The demographic distribution of the 200 participants in our human-subject MIST, depicted in Figure 5, predominantly features younger, highly educated, and left-leaning individuals with a strong trust in scientists. Cognitive features such as trust in scientists, government, and journalism are gauged using questions akin to those designed by Nadelson et al. (2014), a method also employed in other misinformation-related studies (O'Brien et al., 2021). Such a demographic skew is typical in human-subject MISTs, often necessitating large participant pools for balanced representation. The proposed SMIST addresses this challenge by enabling full control over participant demographic groups, allowing for tailored and efficient research under diverse conditions.

When comparing these results to SMIST outcomes with older misinformation (Tables B1 and B2 in Appendix B), *we observed higher mean scores in the information accuracy tests for the newer misinformation across all five topics. Remarkably, these scores closely align with those from the human-subject MISTs,* suggesting that the unfamiliarity of the misinformation prompts the GPT-4 model to rely more heavily on the given participant profiles rather than its own factual knowledge. The OLS regression analyses (Table C1 in Appendix C) further confirm the consistency between SMIST and human-subject MIST results with unfamiliar misinformation. *A notable difference between the two methodologies is observed in the relationship between education level and misinformation susceptibility.* While SMIST suggests negligible or negative correlations between education level and susceptibility to certain misinformation topics, human-subject MIST results indicate consistently positive impacts. *This discrepancy may stem from the skewed educational distribution among our MIST participants, with 67.50% holding a bachelor's degree or higher.* Despite these variances in education level associations, *SMIST findings largely mirror those from our human-subject MISTs for other demographic predictors of misinformation susceptibility.* This congruence underscores SMIST's robustness as a methodology for assessing misinformation susceptibility, regardless of the misinformation's source or recency.

5 Conclusion and Future Work

We introduced the SMIST, incorporating LLM simulations into traditional MIST frameworks for efficient, scalable, and generalizable analyses of misinformation susceptibility across various groups. SMIST offers a cost-effective method for studying factors influencing misinformation susceptibility, aiding the development of psychological inoculations. Our findings validate SMIST's utility in misinformation research, showing alignment with prior studies and human-subject experiments. SMIST employs life stories to simulate participant profiles, addressing the caricature problem common in LLM simulations and fostering diverse responses. This novel approach has the potential for adoption in other simulation-heavy research domains. To further aid the research community, we have made our simulated data and the analytical code publicly available at `https://github.com/hikari-NYU/SMIST`. We encourage researchers to leverage these resources to advance their studies in misinformation and related fields.

Future work could focus on using SMIST for misinformation inoculations, leveraging its capabilities to assess inoculation effectiveness without needing repeated human-subject surveys.

Limitations

While the Simulated Misinformation Susceptibility Test (SMIST) we propose offers significant advances in the field, it is crucial to acknowledge its dependence on the capabilities of LLMs to simulate participant responses in misinformation contexts. SMIST's effectiveness is intertwined with the performance and availability of the underlying LLM (e.g., GPT-4). As a commercial model, GPT-4 may face limitations or alterations in its ability to process queries related to misinformation. Should no comparable models be available or developed, the efficacy of SMIST could be impacted.

However, considering the rapid and transformative advancements in LLM technology, we remain optimistic about these models' future and potential enhancements. Future developments in LLMs could further strengthen the reliability and accuracy of SMIST in emulating human responses in misinformation susceptibility tests. Such progress would be invaluable in advancing research to mitigate the harmful effects of misinformation.

It is also noteworthy that, due to the abundance of misinformation-related research only in English and our need to compare SMIST results with prior studies, we conducted experiments exclusively on English misinformation data, while misinformation-related research is critical in all languages. Since the validity of SMIST has been verified in this paper, future research could leverage SMIST to expand the study of misinformation to other cultures and languages.

Moreover, although SMIST is effective in analyzing the factors influencing misinformation susceptibility and assisting in the design of misinformation inoculations, it remains a sandbox tool and cannot fully replace human participants in these studies. Human experiments are still necessary to confirm the ultimate effectiveness of misinformation mitigation strategies, though intermediate exploratory experiments can be conducted using simulated participants.

Ethics Statement

Our research, encompassing both human-subject and simulation-based experiments using misinformation data, aims to deepen understanding of human vulnerability to misinformation and contribute to its reduction. All misinformation pieces employed in our experiments have been fact-checked and are publicly available rather than being fabricated or disseminated by our team.

While it can be argued that studying misinformation susceptibility with LLMs is a double-edged sword, potentially enabling malicious actors to amplify their efficacy, this complexity actually underscores the critical need for SMISTs. Rapid development of "misinformation inoculations" is essential to counter the swift and targeted spread of misinformation. Reliance solely on human-subject MISTs for developing these inoculations is impractical due to the time-sensitive nature of the problem. Our research aims to proactively position researchers ahead of malicious entities in technological utilization, contributing to this objective. To further reduce misuse of SMISTs, a feasible approach is for owners of these LLMs to allow only trusted partners (e.g., academics) to use their models in the capacity that we have done in this paper.

In our interactions with human participants, we ensured ethical compliance and respect for their well-being. Participants were explicitly informed about the nature of the misinformation involved (related to COVID-19, the Russo-Ukraine war, the 2020 US election, GMOs, and climate change) and were advised against participating if they felt uncomfortable with these topics. They were also explicitly told that their responses to the questionnaires would be used for research purposes.

At the outset of our questionnaire, we also made it clear that participants would be asked to provide demographic information (age, gender, education level, political ideology, trust in scientists, government, and journalism). Importantly, we offered the option to opt out of providing such information. No personally identifiable information was collected to protect participant privacy, and all responses will be de-identified before any public release.

This study's approach to collecting sensitive information has been reviewed and approved by the Institutional Review Board (IRB) at Dartmouth College, ensuring adherence to ethical standards.

Acknowledgements

This work is supported in part by NSF Award 2242072.

References

Lisa P Argyle, Ethan C Busby, Nancy Fulda, Joshua R Gubler, Christopher Rytting, and David Wingate. 2023. Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3):337–351.

K Peren Arin, Deni Mazrekaj, and Marcel Thum. 2023. Ability of detecting and willingness to share fake news. *Scientific Reports*, 13(1):7298.

Melisa Basol, Jon Roozenbeek, and Sander Van der Linden. 2020. Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of cognition*, 3(1).

Dustin P Calvillo, Bryan J Ross, Ryan JB Garcia, Thomas J Smelter, and Abraham M Rutchick. 2020. Political ideology predicts perceptions of the threat of covid-19 (and susceptibility to fake news about it). *Social Psychological and Personality Science*, 11(8):1119–1128.

Dustin P Calvillo, Abraham M Rutchick, and Ryan JB Garcia. 2021. Individual differences in belief in fake news about election fraud after the 2020 us election. *Behavioral Sciences*, 11(12):175.

Myra Cheng, Tiziano Piccardi, and Diyi Yang. 2023. Compost: Characterizing and evaluating caricature in llm simulations. *arXiv preprint arXiv:2310.11501*.

Alton YK Chua and Snehasish Banerjee. 2018. Intentions to trust and share online health rumors: An experiment with medical professionals. *Computers in Human Behavior*, 87:1–9.

David De Coninck, Thomas Frissen, Koen Matthijs, Leen d'Haenens, Grégoire Lits, Olivier Champagne-Poirier, Marie-Eve Carignan, Marc D David, Nathalie Pignard-Cheynel, Sébastien Salerno, et al. 2021. Beliefs in conspiracy theories and misinformation about covid-19: Comparative perspectives on the role of anxiety, depression and exposure to and trust in information sources. *Frontiers in psychology*, 12:646394.

Israel Junior Borges Do Nascimento, Ana Beatriz Pizarro, Jussara M Almeida, Natasha Azzopardi-Muscat, Marcos André Gonçalves, Maria Björklund, and David Novillo-Ortiz. 2022. Infodemics and health misinformation: a systematic review of reviews. *Bulletin of the World Health Organization*, 100(9):544.

Kostas Gemenis. 2021. Explaining conspiracy beliefs and scepticism around the covid-19 pandemic. *Swiss Political Science Review*, 27(2):229–242.

Jiyoung Han, Meeyoung Cha, and Wonjae Lee. 2020. Anger contributes to the spread of covid-19 misinformation. *Harvard Kennedy School Misinformation Review*, 1(3).

Hadas Kotek, Rikker Dockum, and David Sun. 2023. Gender bias and stereotypes in large language models. In *Proceedings of The ACM Collective Intelligence Conference*, pages 12–24.

Kaisheng Lai, Yingxin Yang, Yuxiang Na, and Haixia Wang. 2022. The relationship between bullshit receptivity and willingness to share misinformation about climate change: The moderating role of pregnancy. *International Journal of Environmental Research and Public Health*, 19(24):16670.

Kristin Lunz Trujillo, Matthew Motta, Timothy Callaghan, and Steven Sylvester. 2021. Correcting misperceptions about the mmr vaccine: Using psychological risk factors to inform targeted communication strategies. *Political Research Quarterly*, 74(2):464–478.

Benjamin Lyons, Vittorio Merola, and Jason Reifler. 2019. Not just asking questions: Effects of implicit and explicit conspiracy information about vaccines and genetic modification. *Health communication*, 34(14):1741–1750.

Weicheng Ma, Brian Chiang, Tong Wu, Lili Wang, and Soroush Vosoughi. 2023. Intersectional stereotypes in large language models: Dataset and analysis. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 8589–8597, Singapore. Association for Computational Linguistics.

Louis Nadelson, Cheryl Jorcyk, Dazhi Yang, Mary Jarratt Smith, Sam Matson, Ken Cornell, and Virginia Husting. 2014. I just don't trust them: the development and validation of an assessment instrument to measure trust in science and scientists. *School Science and Mathematics*, 114(2):76–86.

Brendan Nyhan, Jason Reifler, Sean Richey, and Gary L Freed. 2014. Effective messages in vaccine promotion: a randomized trial. *Pediatrics*, 133(4):e835–e842.

Thomas C O'Brien, Ryan Palmer, and Dolores Albarracin. 2021. Misplaced trust: When trust in science fosters belief in pseudoscience and the benefits of critical evaluation. *Journal of Experimental Social Psychology*, 96:104184.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–22.

Kristen Pickles, Erin Cvejic, Brooke Nickel, Tessa Copp, Carissa Bonner, Julie Leask, Julie Ayre, Carys Batcup, Samuel Cornell, Thomas Dakin, et al. 2021. Covid-19 misinformation trends in australia: prospective longitudinal national survey. *Journal of medical Internet research*, 23(1):e23805.

Jon Roozenbeek, Claudia R Schneider, Sarah Dryhurst, John Kerr, Alexandra LJ Freeman, Gabriel Recchia, Anne Marthe Van Der Bles, and Sander Van Der Linden. 2020. Susceptibility to misinformation about covid-19 around the world. *Royal Society open science*, 7(10):201199.

Jon Roozenbeek, Sander Van Der Linden, Beth Goldberg, Steve Rathje, and Stephan Lewandowsky. 2022. Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, 8(34):eabo6254.

Murray Shanahan, Kyle McDonell, and Laria Reynolds. 2023. Role play with large language models. *Nature*, 623(7987):493–498.

Cecilie S Traberg, Jon Roozenbeek, and Sander van der Linden. 2022. Psychological inoculation against misinformation: Current evidence and future directions. *The ANNALS of the American Academy of Political and Social Science*, 700(1):136–151.

Joseph E Uscinski, Adam M Enders, Casey Klofstad, Michelle Seelig, John Funchion, Caleb Everett, Stefan Wuchty, Kamal Premaratne, and Manohar Murthi. 2020. Why do people believe covid-19 conspiracy theories? *Harvard Kennedy School Misinformation Review*, 1(3).

Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, et al. 2023. A survey on large language model based autonomous agents. *arXiv preprint arXiv:2308.11432*.

Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, et al. 2023. The rise and potential of large language model based agents: A survey. *arXiv preprint arXiv:2309.07864*.

Caleb Ziems, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. 2023. Can large language models transform computational social science? *arXiv preprint arXiv:2305.03514*.

A Example of Life Stories

Prompt: Generate the life experiences of 5 elderly people. Please improvise and try to be as diverse as possible on other characteristics of the people when generating the stories.

Figure A1: An example prompt used to generate life experiences for SMIST.

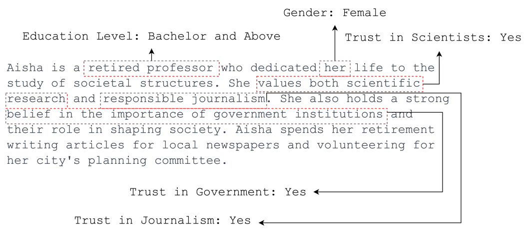

Figure A2: An example life story in our participant pool. The inferred demographic features are also noted.

To enhance diversity in LLM simulations and address the issue of caricatures, we focused on a single demographic feature for each life story generation while allowing the GPT-3.5 model to spontaneously develop other demographic aspects. Manual annotations were subsequently applied to these life stories to identify and categorize the uncontrolled demographic features, facilitating thorough analysis and statistical evaluation. Figure A1 presents an example of the prompts used to generate life stories with "age" set to "elderly", while Figure A2 showcases a generated example, including the manually annotated demographic features.

B Mean and Standard Deviation of SMIST Results

The means and standard deviations of scores retrieved from SMISTs are displayed in Table B1 (for the information accuracy tests) and Table B2 (for the willingness-to-share tests). Most of the standard deviations are low (between 0.10 and 0.60), indicating high in-group robustness of the scores.

C SMIST and Human-Subject MIST Results on Recent Misinformation

Table C1 presents the coefficients obtained from OLS regression analyses of SMIST and human-subject MIST responses. The misinformation stimuli employed in this experimental series were selected based on their recent verification status on Snopes, ensuring their unfamiliarity to both the LLMs and human subjects involved. Consequently, this approach minimizes the potential influence of pre-existing knowledge regarding misinformation among the participants.

		Age			Gender		Education Level			Political Ideology		Trust in Scientists		Trust in Government		Trust in Journalism	
		Young Adult	Middle-Aged	Elderly	Male	Female	Non-Educated	Secondary Education	Bachelor or Above	Liberal	Conservative	Trust	Distrust	Trust	Distrust	Trust	Distrust
COVID-19	Genuine Info	6.80 (0.20)	6.80 (0.20)	7.00 (0.00)	6.80 (0.20)	6.80 (0.20)	6.80 (0.20)	6.80 (0.20)	5.90 (0.40)	6.80 (0.20)	6.20 (0.37)	6.80 (0.20)	6.80 (0.20)	6.67 (0.20)	6.50 (0.20)	6.50 (0.20)	6.67 (0.20)
	Misinfo	2.10 (0.37)	1.60 (0.24)	2.70 (0.51)	1.80 (0.58)	2.10 (0.40)	3.40 (0.24)	2.80 (0.25)	1.50 (0.32)	2.20 (0.20)	2.60 (0.24)	1.20 (0.20)	3.90 (0.78)	1.67 (0.40)	2.25 (0.10)	1.75 (0.10)	1.83 (0.10)
Russo-Ukraine War	Genuine Info	6.40 (0.24)	6.50 (0.32)	6.60 (0.24)	6.60 (0.24)	6.60 (0.24)	6.40 (0.24)	6.80 (0.20)	5.50 (0.22)	6.20 (0.20)	6.40 (0.24)	6.40 (0.24)	6.60 (0.24)	6.67 (0.20)	6.00 (0.15)	6.50 (0.20)	6.67 (0.20)
	Misinfo	3.70 (0.50)	2.60 (0.35)	3.80 (0.59)	2.80 (0.44)	3.40 (0.51)	3.60 (0.33)	3.50 (0.62)	3.10 (0.37)	3.40 (0.37)	3.50 (0.15)	2.40 (0.58)	5.30 (0.32)	3.00 (0.53)	3.25 (0.30)	2.75 (0.40)	3.50 (0.20)
2020 US Election	Genuine Info	6.40 (0.24)	6.60 (0.20)	6.20 (0.20)	6.20 (0.20)	6.80 (0.18)	6.60 (0.30)	6.40 (0.24)	6.50 (0.49)	6.20 (0.20)	6.60 (0.40)	6.60 (0.25)	6.40 (0.24)	6.30 (0.60)	6.80 (0.20)	6.50 (0.20)	6.67 (0.20)
	Misinfo	2.90 (0.43)	2.40 (0.19)	3.30 (0.41)	2.60 (0.62)	3.20 (0.72)	4.00 (0.10)	3.60 (0.29)	2.20 (0.34)	3.60 (0.60)	4.10 (0.78)	2.10 (0.24)	4.00 (0.85)	3.33 (0.80)	2.75 (0.10)	2.50 (0.20)	2.33 (0.10)
GMO	Genuine Info	6.20 (0.20)	6.40 (0.24)	6.20 (0.20)	6.20 (0.25)	6.20 (0.20)	6.60 (0.40)	6.60 (0.40)	6.50 (0.32)	6.40 (0.24)	6.80 (0.17)	6.40 (0.24)	6.20 (0.20)	6.00 (0.40)	6.00 (0.54)	6.00 (0.20)	6.67 (0.20)
	Misinfo	3.70 (0.34)	3.20 (0.49)	4.10 (0.29)	3.60 (0.40)	3.50 (0.32)	4.60 (0.40)	4.20 (0.41)	2.90 (0.40)	4.00 (0.32)	3.40 (0.24)	2.80 (0.37)	3.10 (0.87)	3.67 (0.40)	4.25 (0.10)	3.25 (0.30)	3.50 (0.17)
Climate Change	Genuine Info	6.10 (0.10)	6.10 (0.46)	6.00 (0.00)	6.70 (0.30)	6.20 (0.20)	6.50 (0.32)	6.10 (0.46)	5.40 (0.58)	6.70 (0.30)	6.30 (0.44)	6.30 (0.49)	6.20 (0.20)	6.83 (0.16)	5.25 (0.30)	6.75 (0.25)	6.17 (0.10)
	Misinfo	3.00 (0.55)	2.50 (0.32)	4.00 (0.63)	2.80 (0.72)	2.70 (0.51)	3.70 (0.68)	3.70 (0.46)	2.30 (0.30)	2.50 (0.27)	3.90 (0.37)	2.00 (0.16)	4.40 (0.71)	1.67 (0.10)	3.25 (0.14)	2.75 (0.17)	2.67 (0.10)

Table B1: SMIST Information-Accuracy Results: Mean and Standard Deviation of Raw Scores (Scale 1-7, N=160).

		Age			Gender		Education Level			Political Ideology		Trust in Scientists		Trust in Government		Trust in Journalism	
		Young Adult	Middle-Aged	Elderly	Male	Female	Non-Educated	Secondary Education	Bachelor or Above	Liberal	Conservative	Trust	Distrust	Trust	Distrust	Trust	Distrust
COVID-19	Genuine Info	5.10 (0.40)	6.10 (0.33)	4.20 (0.37)	4.80 (0.20)	4.80 (0.66)	4.40 (0.24)	5.50 (0.39)	5.80 (0.37)	4.80 (0.37)	4.80 (0.20)	6.40 (0.60)	5.00 (0.32)	5.33 (0.53)	4.25 (0.30)	6.00 (0.40)	5.67 (0.20)
	Misinfo	2.00 (0.63)	1.80 (0.25)	2.70 (0.62)	1.80 (0.58)	2.30 (0.30)	2.30 (0.51)	2.70 (0.62)	2.00 (0.27)	2.10 (0.10)	2.30 (0.20)	1.70 (0.24)	5.20 (0.58)	1.50 (0.17)	2.50 (0.20)	1.50 (0.20)	1.33 (0.20)
Russo-Ukraine War	Genuine Info	5.50 (0.24)	6.50 (0.32)	4.50 (0.24)	5.30 (0.24)	5.10 (0.24)	5.10 (0.24)	6.30 (0.20)	5.80 (0.22)	5.30 (0.20)	5.00 (0.24)	6.20 (0.24)	6.00 (0.24)	5.33 (0.20)	4.25 (0.12)	6.00 (0.20)	6.33 (0.20)
	Misinfo	2.40 (0.50)	2.80 (0.35)	2.80 (0.59)	2.60 (0.44)	2.60 (0.51)	2.80 (0.33)	3.40 (0.62)	2.60 (0.37)	2.20 (0.37)	2.90 (0.10)	2.40 (0.58)	2.20 (0.32)	3.50 (0.53)	2.00 (0.30)	3.00 (0.40)	2.33 (0.20)
2020 US Election	Genuine Info	6.10 (0.84)	6.50 (0.37)	5.60 (0.51)	5.00 (0.32)	6.00 (0.32)	5.20 (0.58)	6.50 (0.39)	6.20 (0.66)	6.40 (0.60)	6.00 (0.55)	6.40 (0.24)	6.60 (0.35)	5.67 (0.20)	4.75 (0.50)	6.00 (0.17)	6.33 (0.40)
	Misinfo	2.00 (0.27)	1.80 (0.12)	2.10 (0.24)	2.00 (0.27)	1.90 (0.10)	1.70 (0.12)	3.90 (1.10)	2.10 (0.10)	2.00 (0.34)	2.00 (0.27)	2.00 (0.27)	1.90 (0.10)	2.50 (0.17)	1.75 (0.10)	1.75 (0.20)	1.83 (0.17)
GMO	Genuine Info	5.90 (0.51)	6.50 (0.45)	5.40 (0.68)	5.00 (0.32)	5.40 (0.75)	4.80 (0.37)	6.30 (0.54)	6.60 (0.32)	6.20 (0.58)	6.20 (0.49)	6.00 (0.55)	6.40 (0.51)	6.00 (0.35)	5.25 (0.70)	6.50 (0.20)	6.33 (0.20)
	Misinfo	3.00 (0.32)	2.60 (0.24)	3.00 (0.32)	2.80 (0.20)	2.60 (0.24)	3.00 (0.33)	3.60 (0.24)	2.40 (0.24)	2.40 (0.24)	2.90 (0.40)	2.60 (0.40)	2.60 (0.24)	2.83 (0.36)	2.50 (0.20)	3.00 (0.21)	2.67 (0.20)
Climate Change	Genuine Info	5.40 (0.91)	5.80 (0.77)	4.20 (0.77)	4.80 (0.58)	4.90 (0.89)	5.30 (0.41)	5.80 (0.73)	5.30 (0.78)	5.40 (0.33)	4.50 (0.63)	6.20 (0.58)	5.00 (0.92)	6.00 (0.17)	3.50 (0.60)	6.75 (0.10)	6.17 (0.80)
	Misinfo	2.10 (0.24)	1.90 (0.10)	2.00 (0.27)	2.00 (0.27)	1.70 (0.12)	2.00 (0.24)	3.20 (0.77)	2.00 (0.16)	1.80 (0.12)	2.00 (0.27)	1.90 (0.29)	1.90 (0.10)	2.33 (0.26)	1.50 (0.42)	2.00 (0.25)	1.67 (0.10)

Table B2: SMIST Willingness-to-Share Results: Mean and Standard Deviation of Raw Scores (Scale 1-7, N=160).

Topic	Approach	Questionnaire Type	Intercept	Age	Gender (female)	Education Level	Political Ideology (conservative)	Trust in Scientists	Trust in Government	Trust in Journalism
COVID-19	SMIST	accuracy	2.56 [2.53,2.58]	-0.69 [-0.86,0.52]	-0.60 [-0.76,0.44]	0.00 [-0.05,0.04]	0.11 [0.09,0.13]	-0.04 [-0.05,-0.02]	-0.09 [-0.20,0.01]	-0.11 [-0.19,-0.03]
		willingness-to-share	4.35 [4.34,4.36]	-0.32 [-0.50,-0.13]	0.00 [-0.18,0.19]	-0.02 [-0.03,-0.01]	0.03 [-0.07,0.13]	-0.07 [-0.12,-0.03]	-0.28 [-0.53,-0.02]	0.14 [0.08,0.20]
	MIST	accuracy	3.67 [3.52,3.83]	-0.17 [-0.21,-0.13]	-0.19 [-0.23,-0.15]	0.23 [0.19,0.26]	0.11 [0.07,0.15]	-0.48 [-0.54,-0.43]	-0.05 [-0.09,-0.01]	-0.32 [-0.36,-0.29]
		willingness-to-share	3.40 [3.24,3.55]	-0.12 [-0.15,-0.09]	-0.27 [-0.31,-0.23]	0.28 [0.26,0.30]	0.08 [0.04,0.11]	-0.23 [-0.28,-0.18]	-0.17 [-0.20,-0.14]	-0.08 [-0.11,-0.04]
Russo-Ukraine War	SMIST	accuracy	4.36 [4.34,4.38]	0.12 [0.02,0.22]	0.14 [0.04,0.24]	-0.14 [-0.27,0.00]	0.03 [0.02,0.03]	-0.10 [-0.20,0.00]	0.11 [-0.01,0.24]	-0.14 [-0.18,-0.10]
		willingness-to-share	2.70 [2.68,2.72]	0.03 [-0.10,0.16]	-0.09 [-0.12,-0.06]	0.26 [0.16,0.36]	0.07 [0.01,0.14]	-0.10 [-0.14,-0.06]	0.05 [-0.01,0.11]	-0.08 [-0.15,0.00]
	MIST	accuracy	2.93 [2.77,3.09]	0.01 [-0.03,0.04]	0.06 [0.03,0.10]	0.50 [0.47,0.53]	0.03 [0.00,0.07]	-0.34 [-0.39,-0.28]	0.19 [0.15,0.22]	-0.33 [-0.37,-0.29]
		willingness-to-share	2.63 [2.48,2.78]	-0.13 [-0.16,-0.10]	-0.23 [-0.26,-0.19]	0.40 [0.37,0.43]	-0.09 [-0.12,-0.06]	-0.12 [-0.17,-0.07]	0.19 [0.15,0.22]	-0.03 [-0.07,0.00]
2020 US Election	SMIST	accuracy	2.62 [2.58,2.65]	0.11 [0.04,0.18]	0.32 [0.14,0.51]	-0.29 [-0.38,-0.20]	-0.07 [-0.13,-0.01]	-0.02 [-0.04,0.00]	0.01 [-0.01,0.02]	-0.14 [-0.24,-0.05]
		willingness-to-share	2.34 [2.32,2.37]	-0.06 [-0.22,0.11]	0.21 [0.15,0.28]	0.06 [0.00,0.12]	-0.07 [-0.11,-0.02]	-0.07 [-0.12,-0.03]	-0.04 [-0.07,-0.01]	-0.11 [-0.20,-0.02]
	MIST	accuracy	3.01 [2.87,3.16]	0.04 [0.01,0.07]	0.04 [0.01,0.08]	0.46 [0.43,0.48]	0.04 [0.01,0.07]	-0.27 [-0.32,-0.23]	0.03 [0.01,0.05]	-0.29 [-0.32,-0.26]
		willingness-to-share	2.62 [2.46,2.77]	-0.01 [-0.04,0.02]	-0.41 [-0.45,-0.37]	0.3 [0.27,0.33]	-0.13 [-0.16,-0.10]	-0.26 [-0.32,-0.21]	-0.1 [-0.14,-0.06]	-0.02 [-0.06,0.01]
GMO	SMIST	accuracy	2.74 [2.70,2.77]	0.41 [0.14,0.68]	0.31 [0.21,0.41]	0.37 [0.22,0.52]	0.06 [0.00,0.11]	-0.03 [-0.05,-0.01]	0.23 [0.22,0.23]	-0.21 [-0.24,-0.17]
		willingness-to-share	2.39 [2.35,2.43]	0.11 [-0.03,0.26]	0.23 [0.19,0.28]	0.16 [0.13,0.20]	0.08 [0.02,0.14]	-0.03 [-0.05,-0.01]	0.10 [0.02,0.19]	-0.02 [-0.02,-0.02]
	MIST	accuracy	3.57 [3.42,3.71]	0.05 [0.02,0.08]	0.27 [0.24,0.30]	0.36 [0.33,0.38]	0.05 [0.02,0.08]	-0.69 [-0.74,-0.64]	0.17 [0.13,0.20]	-0.38 [-0.41,-0.35]
		willingness-to-share	2.32 [2.19,2.46]	-0.07 [-0.10,-0.04]	0.34 [0.31,-0.37]	0.19 [0.16,0.21]	0.00 [-0.03,0.03]	-0.31 [-0.36,-0.26]	0.20 [0.17,0.23]	-0.05 [-0.08,-0.02]
Climate Change	SMIST	accuracy	2.35 [2.33,2.37]	0.32 [0.16,0.47]	0.16 [0.00,0.31]	0.38 [0.21,0.55]	-0.04 [-0.06,-0.01]	0.21 [0.19,0.23]	-0.05 [-0.06,-0.03]	0.36 [0.36,0.37]
		willingness-to-share	4.00 [3.99,4.02]	0.04 [-0.05,0.14]	0.09 [0.00,0.19]	0.34 [0.11,0.57]	-0.03 [-0.06,-0.01]	0.18 [0.10,0.27]	-0.19 [-0.36,-0.03]	0.21 [0.05,0.38]
	MIST	accuracy	3.12 [2.96,3.27]	0.02 [-0.01,0.05]	0.11 [0.07,-0.15]	0.27 [0.24,0.30]	-0.03 [-0.06,0.00]	0.58 [0.53,0.63]	-0.30 [-0.34,-0.26]	0.29 [0.26,0.32]
		willingness-to-share	2.41 [2.26,2.55]	0.13 [0.10,0.16]	0.21 [0.17,0.25]	0.22 [0.19,0.24]	-0.02 [-0.05,0.01]	0.46 [0.41,0.51]	-0.25 [-0.29,-0.21]	0.02 [0.01,0.03]

Table C1: SMIST and Human-Subject MIST OLS Regression Coefficients on Recent Misinformation (95% Confidence Intervals in Brackets).