Agent-Based Analysis of Personality Trait Effects on Misinformation Dissemination

Methodology

Agent Design and Personality Assignment

Six different agents were used in this study, each representing one of the key personality traits of Extraversion, Agreeableness, and Neuroticism based on the Big Five personality traits. The agents were assigned as follows:

Agent 1	extraversion(bold/energetic)
Agent 2	extraversion(shy/bashful)
Agent 3	agreeableness (sympathetic/cooperative)
Agent 4	agreeableness (cold/harsh)
Agent 5	neuroticism(moody/nervous)
Agent 6	neuroticism(relaxed/calm)

Interaction Scenarios and Outcomes

Any pair of agents engaged in a conversation focusing on a specific misinformation topic. The interaction resulted in one of the following four possible outcomes:

- Agent A convinces Agent B: Agent A successfully persuades Agent B to adopt its position.
- Agent B convinces Agent A: Agent B successfully persuades Agent A to adopt its position.
- Mutual Resistance: Both agents maintain their original beliefs.
- Bilateral Influence: Two agents influence each other, resulting in both Agent A persuading Agent B and Agent B persuading Agent A in the conversation.

Misinformation Topics and Conversational Frames

Six different misinformation topics were selected, representing different categories such as health myths, conspiracy theories, and technology misconceptions. These include:

- Misconceptions about the safety of the MMR vaccine.
- False claims about the causes of HIV disease.
- Conspiracy theories about the advancement of 5G technology.
- Misleading dietary advice about superfoods.
- Misinformation about chloride causing certain diseases.
- Fabricated QAnon theory (deep state theory).

Simulation Framework

The experiment was implemented using the AgentScope framework, which facilitates multi-agent interactions with customizable personality and behavior settings.

Experimental Design and Data Collection

To analyze the impact of personality traits on misinformation dynamics, all six agents engaged in pairwise discussions on each of the six misinformation topics. This resulted in a total of 15 unique agent pair combinations per topic and 90 interactions across all topics. For each interaction, the number of times the four scenarios occurred and their frequency were recorded.

Model Selection and Implementation

The study utilized a large language model (GLM-4-Flash) to simulate agent behavior and generate conversation responses. The LLM was fine-tuned to reflect the specified personality traits of each agent, ensuring realistic and consistent interactions. Agents' decisions are influenced by their personality profiles, allowing exploration of how these traits affect their sensitivity or resistance to misinformation.

Ethics checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes

(d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? No

(e) Did you describe the limitations of your work? Yes

(f) Did you discuss any potential negative societal impacts of your work? Yes

(g) Did you discuss any potential misuse of your work? No

(h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes

(i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes

2. Additionally, if your study involves hypotheses testing...

(a) Did you clearly state the assumptions underlying all theoretical results? NA

(b) Have you provided justifications for all theoretical results? NA

(c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA

(d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA

(e) Did you address potential biases or limitations in your theoretical framework? NA

(f) Have you related your theoretical results to the existing literature in social science? NA

(g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA

3. Additionally, if you are including theoretical proofs...

(a) Did you state the full set of assumptions of all theoretical results? NA

(b) Did you include complete proofs of all theoretical results? NA

4. Additionally, if you ran machine learning experiments...

(a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? No

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? No

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? No

(e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? No

(f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? No

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...

(a) If your work uses existing assets, did you cite the creators? NA

(b) Did you mention the license of the assets? NA

(c) Did you include any new assets in the supplemental material or as a URL? NA

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? NA

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see (**?**))? NA

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see (**?**))? NA

6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...

(a) Did you include the full text of instructions given to participants and screenshots? NA

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA

(d) Did you discuss how data is stored, shared, and deidentified? NA