

Simulating Influence Dynamics with LLM Agents

Mehwish Nasim¹, Syed Muslim Gilani¹, Amin Qasmi² and Usman Naseem³

¹The University of Western Australia

²Lahore University of Management Sciences

³Macquarie University

{mehwish.nasim, syedmuslim.gilani}@uwa.edu.au, amin.qasmi@lums.edu.pk,
usman.naseem@mq.edu.au

Abstract

This paper introduces a simulator designed for opinion dynamics researchers to model competing influences within social networks in the presence of LLM-based agents. By integrating established opinion dynamics principles with state-of-the-art LLMs, this tool enables the study of influence propagation and counter-misinformation strategies. The simulator is particularly valuable for researchers in social science, psychology, and operations research, allowing them to analyse societal phenomena without requiring extensive coding expertise. Additionally, the simulator will be openly available on GitHub, ensuring accessibility and adaptability for those who wish to extend its capabilities for their own research.

1 Introduction

Large Language Models (LLMs) are becoming ubiquitous, often shaping discourse in ways we barely notice. But what happens when the entire public opinion space is influenced or even outsourced to AI-driven agents [Yu, 2024]? While LLMs have been extensively studied in isolation, their behavior within dynamic social networks, interacting alongside humans, remains an open and critical research frontier. Understanding how these AI-enabled agents shape influence, polarisation, and consensus in evolving networks is key to anticipating the societal impacts of this technological shift. [Kudrator, 2024], [Zhao *et al.*, 2024], [Leng and Yuan, 2023], [Papachristou and Yuan, 2024].

Understanding how people adjust their opinions based on social influence was the basis of opinion dynamics research [Kelman, 1958], [Kelman, 1961], with wide-ranging implications in fields such as public health initiatives, conflict resolution, and misinformation mitigation. Opinions spread and evolve within social networks, often driven by factors such as peer influence [Kandel, 1986], media exposure [Zucker, 1978], and group dynamics [Friedkin and Johnsen, 2011]. Accurate models of these processes have been considered critical not only for forecasting trends such as opinion polarisation [Tan *et al.*, 2024] or consensus formation, but also for designing targeted interventions to counteract harmful effects, such as the spread of misinformation or societal di-

vides [Hegselmann and Krause, 2015]. Agent-based models (ABMs) are used to simulate interactions among individual agents (a proxy for humans) to explore the emergent properties of opinion propagation. They can provide powerful frameworks for investigating complex scenarios [Deffuant *et al.*, 2002], [Mathias *et al.*, 2016], for testing strategies for mitigating negative outcomes and perhaps fostering constructive social influence, e.g., incorporating explicit assumptions about cognitive processes in opinion updating.

Understanding how LLMs behave in multi-agent social interactions is crucial for advancing AI applications [Tang *et al.*, 2024], [Lan *et al.*, 2024]. LLMs in autonomous systems offer opportunities to revolutionise decision-making by simulating fairness, reciprocity, and competition in social contexts [Wang *et al.*, 2024]. Their behaviour could influence resource allocation, conflict resolution, and interaction strategies. Unlike traditional agent-based models with pre-defined rules, LLMs can exhibit more flexible, human-like behaviours, enhancing realism in simulations for policy evaluation. These capabilities make them valuable for designing AI systems that better mimic human social dynamics [Horton, 2023], improving both their practical application and the insights they provide into complex, real-world decision-making processes.

This paper introduces a simulator to model influence and counter-influence in a wargame setting. Wargames, originally developed for military strategy, have evolved into powerful tools for decision-making across various domains. Today, they are used to model business strategies, assess cybersecurity threats, and simulate geopolitical conflicts. Governments and corporations employ wargames to anticipate economic shifts, supply chain disruptions, and the impact of emerging technologies. In healthcare, they help model pandemic responses, testing different policy interventions before real-world implementation. AI-driven wargames further enhance scenario analysis, enabling rapid adaptation to complex environments. By fostering strategic thinking and resilience, modern wargaming serves as a critical tool for navigating uncertainty in an increasingly interconnected world.

The simulator can facilitate studies to understand how artificial intelligence, specifically LLMs, can emulate human-like opinion dynamics and influence propagation in a social network. Traditional approaches to modelling opinion dynamics often rely on simplified rules that may not cap-

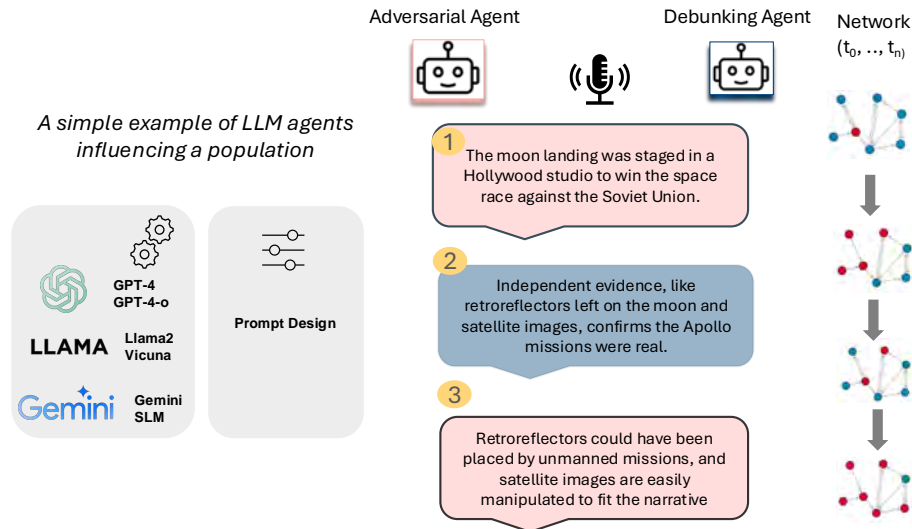


Figure 1: Architecture of the model. Each agent (red/blue), broadcast a message with a potency to affect the population. The nodes in the network receive those messages. They also interact with their direct neighbors. During the interaction they may change their opinion.

ture some of the communicative strategies and adaptive behaviours seen in human interactions. The specific problem tackled by this work is the challenge of understanding the interplay between misinformation and counter-misinformation in shaping public opinion. By introducing adversarial LLMs based agents, for instance, one agent spreading misinformation and the other countering it, this introduces a more realistic framework for analysing how LLMs dominate each other while aiming at shifting the opinion of the population [Chen *et al.*, 2024], [Qu and Wang, 2024], [Aher *et al.*, 2023], [Carpentras, 2023], [Flache *et al.*, 2017].

2 Scenario

The scenario has been strategically developed to reflect the asymmetric nature of the contested information environment, emphasising the vulnerabilities faced by the Blue team. This framework mirrors adversarial dynamics often modeled in serious games or wargames, particularly in cybersecurity. While the Red Team and Blue Team construct is common in cybersecurity practices (as detailed in NIST’s Glossary [National Institute of Standards and Technology (NIST), 2015]), this scenario extends the concept to the broader geopolitical information landscape within a fictitious nation-state [Nasim *et al.*, 2022].

The system comprises two LLM-based agents with opposing objectives: the *Red Agent*, responsible for disseminating misinformation, and the *Blue Agent*, tasked with counteracting misinformation and restoring trust. These agents operate within a directed network of neutral agents, termed *Green Nodes*, which represent individuals within a population. The simulator allows the users to upload their own graphs or use the functionality provided in the simulator to generate a network. Users can choose the LLMs for both Blue and Red agents. Currently, the simulator supports various versions of Open AI’s GPT [OpenAI, 2023] as well as other open source

models from HuggingFace. The simulator also has the provision to upload a new model.

Green nodes exhibit predispositions toward either agent, influenced by prior interactions and the content of incoming messages. Each Green Node’s behaviour is defined by core parameters adapted from the Deffuant model [Deffuant *et al.*, 2002], [Mathias *et al.*, 2016], including susceptibility to influence, confirmation bias, and mechanisms for updating beliefs. These parameters ensure that the modelled population exhibits realistic characteristics, such as resistance to extreme viewpoints and gradual alignment shifts. Each agent in the population is represented by a scalar value (or vector) that denotes their opinion on a specific topic (Figure 2). The opinions are within a bounded range, such as $[0, 1]$. If the difference in their opinions is below a certain threshold (the confidence bound, ϵ), the agents influence each other and adjust their opinions closer together. The adjustment is controlled by a convergence parameter (μ), dictating how much the agents move toward each other’s opinions.

2.1 Simulation Dynamics

The simulation proceeds in discrete time steps, during which the Red/Blue agents alternately broadcast messages to the Green Nodes (also viewable by the other LLM agent). The Green nodes that are connected to each other interact with each other (Figure 1). Key operational components include:

Message Generation Each agent generates a message based on its LLM’s output, informed by the current state of the network and its strategic objective. For example, the Red Agent prioritises persuasive misinformation, while the Blue Agent constructs factual rebuttals optimised for resource efficiency.

Message Potency/influence factor Messages are assigned a potency score that quantifies their influence. The LLMs determine the potency of each message that they generate. The influence factor determines the extent to which

the Green Nodes adjust their alignment toward the broadcasting agent. While the Red Agent has access to unlimited resources, high-potency messages incur penalties, particularly when directed at strongly blue-aligned nodes, mimicking real-world scepticism toward overt misinformation [Ecker *et al.*, 2022]. In contrast, the Blue Agent operates under constrained resources, with each message incurring a cost proportional to its potency. This constraint requires strategic resource management, as overly powerful debunking messages risk rapid depletion of available energy.

Node Update Mechanism Upon receiving a message, the Green Nodes adjust their alignment based on their predisposition, the potency of the message, and the influence of the connected neighbours. Updates occur iteratively, capturing both direct and network-mediated effects of influence propagation.

Termination Criteria The simulation concludes when an agent achieves a majority alignment within the Green Node population, indicating a decisive shift in opinion. Alternatively, the simulation terminates after a fixed number of rounds if neither agent achieves dominance, representing a stalemate.

Confirm Parameter Selections:

Blue Agent	Red Agent	Population	End conditions
Model: gpt-4o-turbo Alignment: 50% Energy Level: 100 Influence Factor: 0.6 Number of Messages Generated Per Turn: 5 Temperature: 1 Maximum Cost: 13	Model: gpt-4o-turbo Alignment: 50 % Influence Factor: 0.5 Number of Messages Generated Per Turn: 5 Temperature: 1 Penalise Messages with Potency of: 50	Network Size: 40 Blue Aligned Nodes: 20 Red Aligned Nodes: 20 Neutral Nodes: 0	Population alignment: 80% Round number: 30

Enter a Simulation Topic (Optional):

A secret organisation controls the government

Once confirmed these settings are correct, select how you want to play and click "Start Simulation"

Play Continuously
Play in Turns

Start Simulation

Figure 2: User is prompted to confirm the settings and enter a topic.

2.2 Evaluation

The simulation can be evaluated using the following metrics. At the end of the simulation, a .csv file is generated which can be used for further analysis. In addition, messages and network states are also captured.

- **Network Alignment Distribution:** The final proportion of Green Nodes aligned with each agent. This refers to the polarisation in the network. A sample output graph is shown in Figure 3.

- **Resource Efficiency:** The Blue Agent’s energy expenditure relative to alignment gains.
- **Node Resilience:** The resistance of nodes with strong predispositions to opposing influences.
- **Temporal Evolution:** The rate of alignment change over successive rounds.

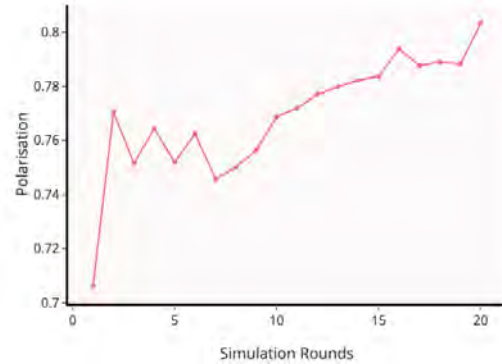


Figure 3: Polarisation in the network over time.

3 Conclusions and Future Work

The simulator presented in this paper provides an interesting approach to studying opinion dynamics, combining the generative capabilities of LLMs with structured agent-based modeling principles. By incorporating realistic constraints, such as resource limitations and susceptibility penalties, it offers insights into the dynamics of influence competition and the effectiveness of counter-misinformation strategies. Furthermore, this work highlights the dual potential of LLMs as both tools for studying opinion propagation and as models for emulating human-like decision-making in complex social systems. We are working on improving the prompting strategies and providing more control to the end user in future.

4 Ethics Statement

We have avoided sharing detailed prompts in the code to prevent misuse. We commit to promoting responsible AI development.

Acknowledgment

This research was supported by the Collaborative Research Grant awarded to Mehwish Nasim by DSC/JTISI Western Australia in 2023. The authors acknowledge the support of the following students in implementing this software: Rhiana Hepburn, JJ Jun, Olivia Morrison, Devarsh Patel and Edwin Tang.

References

- [Aher *et al.*, 2023] Gati V Aher, Rosa I Arriaga, and Adam Tauman Kalai. Using large language models to simulate multiple humans and replicate human subject studies. In *International Conference on Machine Learning*, pages 337–371. PMLR, 2023.

- [Carpentras, 2023] Dino Carpentras. Why we are failing at connecting opinion dynamics to the empirical world. *Review of Artificial Societies and Social Simulations*, 2023.
- [Chen *et al.*, 2024] Kai Chen, Zihao He, Jun Yan, Taiwei Shi, and Kristina Lerman. How susceptible are large language models to ideological manipulation? *arXiv preprint arXiv:2402.11725*, 2024.
- [Deffuant *et al.*, 2002] Guillaume Deffuant, Frédéric Amblard, Gérard Weisbuch, and Thierry Faure. How can extremism prevail? a study based on the relative agreement interaction model. *Journal of artificial societies and social simulation*, 5(4), 2002.
- [Ecker *et al.*, 2022] Ullrich KH Ecker, Stephan Lewandowsky, John Cook, Philipp Schmid, Lisa K Fazio, Nadia Brashier, Panayiota Kendeou, Emily K Vraga, and Michelle A Amazeen. The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1):13–29, 2022.
- [Flache *et al.*, 2017] Andreas Flache, Michael Mäs, Thomas Feliciani, Edmund Chattoe-Brown, Guillaume Deffuant, Sylvie Huet, and Jan Lorenz. Models of social influence: Towards the next frontiers. *Jasss-The journal of artificial societies and social simulation*, 20(4):2, 2017.
- [Friedkin and Johnsen, 2011] Noah E Friedkin and Eugene C Johnsen. *Social influence network theory: A sociological examination of small group dynamics*, volume 33. Cambridge University Press, 2011.
- [Hegselmann and Krause, 2015] Rainer Hegselmann and Ulrich Krause. Opinion dynamics under the influence of radical groups, charismatic leaders, and other constant signals: A simple unifying model. *Networks and Heterogeneous Media*, 10(3):477–509, 2015.
- [Horton, 2023] John J Horton. Large language models as simulated economic agents: What can we learn from homo silicus? Technical report, National Bureau of Economic Research, 2023.
- [Kandel, 1986] Denise B Kandel. Processes of peer influences in adolescence. In *Development as action in context: Problem behavior and normal youth development*, pages 203–227. Springer, 1986.
- [Kelman, 1958] Herbert C Kelman. Compliance, identification, and internalization three processes of attitude change. *Journal of conflict resolution*, 2(1):51–60, 1958.
- [Kelman, 1961] Herbert C Kelman. American association for public opinion research. *The Public Opinion Quarterly*, 25(1):57–78, 1961.
- [Kudiabor, 2024] Helena Kudiabor. Virtual lab powered by ‘AI scientists’ super-charges biomedical research [preprint]. *Nature*, December 2024.
- [Lan *et al.*, 2024] Yihuai Lan, Zhiqiang Hu, Lei Wang, Yang Wang, Deheng Ye, Peilin Zhao, Ee-Peng Lim, Hui Xiong, and Hao Wang. LLM-based agent society investigation: Collaboration and confrontation in avalon gameplay. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 128–145, Miami, Florida, USA, November 2024. Association for Computational Linguistics.
- [Leng and Yuan, 2023] Yan Leng and Yuan Yuan. Do llm agents exhibit social behavior? *arXiv preprint arXiv:2312.15198*, 2023.
- [Mathias *et al.*, 2016] Jean-Denis Mathias, Sylvie Huet, and Guillaume Deffuant. Bounded confidence model with fixed uncertainties and extremists: The opinions can keep fluctuating indefinitely. *Journal of Artificial Societies and Social Simulation*, 19(1):6, 2016.
- [Nasim *et al.*, 2022] Mehwish Nasim, Richard Stearne, and Tim Legrand. A framework for modelling social influence in a wargame setting, May 2022. Completion Report Submitted to the Funding Agency.
- [National Institute of Standards and Technology (NIST), 2015] National Institute of Standards and Technology (NIST). Red team, 2015.
- [OpenAI, 2023] OpenAI. Gpt-4 technical report. <https://openai.com/research/gpt-4>, 2023. Accessed: 2024-12-19.
- [Papachristou and Yuan, 2024] Marios Papachristou and Yuan Yuan. Network formation and dynamics among multi-llms. *arXiv preprint arXiv:2402.10659*, 2024.
- [Qu and Wang, 2024] Yao Qu and Jue Wang. Performance and biases of large language models in public opinion simulation. *Humanities and Social Sciences Communications*, 11(1):1–13, 2024.
- [Tan *et al.*, 2024] Eugene Tan, Thomas Stemler, and Michael Small. Cognitive dissonance and introversion effects on opinion dynamics and echo chamber formation. *Physica A: Statistical Mechanics and its Applications*, 2024.
- [Tang *et al.*, 2024] Jiakai Tang, Heyang Gao, Xuchen Pan, Lei Wang, Haoran Tan, Dawei Gao, Yushuo Chen, Xu Chen, Yankai Lin, Yaliang Li, et al. Gensim: A general social simulation platform with large language model based agents. *arXiv preprint arXiv:2410.04360*, 2024.
- [Wang *et al.*, 2024] Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, et al. A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6):186345, 2024.
- [Yu, 2024] Chen Yu. How will ai steal our elections? *Center for Open Science*, 2024.
- [Zhao *et al.*, 2024] Andrew Zhao, Daniel Huang, Quentin Xu, Matthieu Lin, Yong-Jin Liu, and Gao Huang. Expel: Llm agents are experiential learners. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19632–19642, 2024.
- [Zucker, 1978] Harold Gene Zucker. The variable nature of news media influence. *Annals of the International Communication Association*, 2(1):225–240, 1978.