

Synthetic Social Media Influence Experimentation via an Agentic Reinforcement Learning Large Language Model Bot

Bailu Jin¹ and Weisi Guo¹

¹Cranfield University, College Rd, Wharley End, Bedford MK43 0AL, United Kingdom
Correspondence should be addressed to weisi.guo@cranfield.ac.uk

Journal of Artificial Societies and Social Simulation 28(3) 6, 2025
Doi: 10.18564/jasss.5726 Url: <http://jasss.soc.surrey.ac.uk/28/3/6.html>

Received: 29-11-2024 Accepted: 29-05-2025 Published: 30-06-2025

Abstract: Understanding the dynamics of public opinion evolution on online social platforms is crucial for understanding influence mechanisms and the provenance of information. Traditional influence analysis is typically divided into qualitative assessments of personal attributes (e.g., psychology of influence) and quantitative evaluations of influence power mechanisms (e.g., social network analysis). One challenge faced by researchers is the ethics of real-world experimentation and the lack of social influence data. In this study, we provide a novel simulated environment that combines agentic intelligence with Large Language Models (LLMs) to test topic-specific influence mechanisms ethically. Our framework contains agents that generate posts, form opinions on specific topics, and socially follow/unfollow each other based on the outcome of discussions. This simulation allows researchers to observe the evolution of how opinions form and how influence leaders emerge. Using our own framework, we design an opinion leader that utilizes Reinforcement Learning (RL) to adapt its linguistic interaction with the community to maximize its influence and followers over time. Our current findings reveal that constraining the action space and incorporating self-observation are key factors for achieving stable and consistent opinion leader generation for topic-specific influence. This demonstrates the simulation framework's capacity to create agents that can adapt to complex and unpredictable social dynamics. The work is important in an age of increasing online influence on social attitudes and emerging technologies.

Keywords: Agent-Based Modeling, Agentic Large Language Model, Reinforcement Learning

● Introduction

- 1.1 The emergence of online social media has revolutionized the way we share information and conduct debate. In order to answer key social research questions, we often have to create social influence simulation engines that involve modeling agents (how they behave and learn), how the agents socially network together, and how they disseminate information to each other.
- 1.2 Generally, Agent-Based Modeling (ABM) have proven effective in analyzing and explaining social phenomena by incorporating personal and environmental details. ABMs can use expert-driven models or data-driven learning functions to capture the nonlinear societal mechanisms and reveal large-scale societal emergent effects (Jackson et al. 2017). Within agent modeling, reinforcement learning (RL) has the capability to mimic the human-like action selection. Using RL model, the target agent learn optimal strategies based on the interaction with social environment. These models have been developed to capture our understanding of a social environment that involves multiple agents interacting through language and physical encounters. The social observations that often interest us include: social consensus, clustering, and bi-polarity in shaping public opinion. ABMs also allow computation social science researchers to test hypothesis in more ethical environments without direct real world implications, and create useful data sets that are often lacking from the real world. A current limitation is that ABMs primarily function through mathematical representations and do not inherently integrate natural

language behaviour. Emerging research, which we review in Section II will focus on developing Large Language Models (LLMs) that are controlled by RL agents to ingest and produce natural language interactions with other agents.

Research problem statement and aim

- 1.3 There remain gaps between traditional small-scale psychological analysis of peer influence and the macroscopic analysis of large-scale social media influence. These gaps include the areas of intricacy, discontinuity, and heterogeneity in individual behaviors. Our research aim to bridge this gap between computational simulations and linguistic data by integrating RL-agentic intelligence and Large Language Models (LLMs).
- 1.4 In our research, we have developed a simulated environment that replicates the dynamics of group discussions on a specific topic in an online social media context. This environment is structured with agents representing individual participant and follow relationships as links. Agents release topic-related posts to represent the current opinion and the links will be updated due to opinion states. LLMs are used for post generation and opinion elicitation for agents. The generated posts are not only relevant to the topic but also reflect the complex influences present in a social media environment. The simulation result aligns with the polarity distribution of the actual dataset, indicating that our simulated environment reflects realistic discussion patterns.

Novelty and contributions

- 1.5 In the paper, we specifically focus on training RL models on a target agent with the objective of maximizing its follower count in the simulated network as a proxy measure for opinion shaping. Our study explores various environment and observability settings to understand their impact on the agent's ability to achieve this goal. In our designed cases, the convergence in RL reward learning curves indicating RL achieved optimal solutions.
- 1.6 Here we summarize the major contributions of this paper:
 - Developing a simulation framework on online social networks that integrate natural language interactions through LLMs.
 - Integrating agents that can make decisions on how to improve its influence using reinforcement learning.
 - Our research findings show that limiting the action space of the agent and incorporating self-observation are key factors for stable generation of opinion leaders.
- 1.7 The potential applications of this research are diverse across multiple domains. Influence models can allow us to better understand where misinformation has come from and why certain sources are more successful than others. This affects both policy makers as well as homeland security. For instance, social acceptance of emerging technologies like nanotechnologies and future 6G telecommunications is crucial for business and government (Briguglio et al. 2021), policymakers could use the tool to simulate and evaluate the public's response to communication strategies. In business intelligence, marketing firms could leverage the model to predict the effectiveness of promotional strategies by simulating how opinions evolve in response to targeted campaigns.

● Related Work

Opinion leader

- 2.1 As discussed in Section 1, we first review existing work in ABM and how RL is a particular way for the agent to learn from data on achieving a desired social effect (e.g., opinion influence) in an environment. Here we review detailed works in ABMs, RL models, and language model interaction environments related to social opinion dynamics.

Agent-based modelling (ABM)

- 2.2 Agent-based modelling has been widely used by social researchers in the field of social simulation. In the modelling, there are generally two elements: (1) how agents interact with each other through the exchange of information, and (2) how agents process any information and adapt their behaviour and/or opinion of a topic.

Compared to the other approaches, such as laboratory or field experiments, ABMs apply the experimental control on a large-scale population to capture the nonlinear societal mechanisms and reveal large-scale societal emergence (Jackson et al. 2017). However, high control of ABMs leads to a low degree of external validity. To solve this problem, Betz presented a natural-language agent-based model of argumentation (ABMA) to simulate the argumentative opinion dynamics (Betz 2022). The explicit stance used in traditional simulation of opinion dynamics can be represented by implicit argument from natural-language ABMAs. A number of recent works have used ABMs to integrate generative language models (GABMs) to create social networks.

2.3 One approach to model realistic agent interactions is to simulate the life of agents in a sandbox and information exchange occurs through physical encounters (Park et al. 2023). Physical interactions aid the spread of information, and this has been expanded to model diverse demographics (Park et al. 2023). To reduce the complexity of simulations, other models adopt a norm-diffusion model via LLMs (Ghaffarzadegan et al. 2024). Our approach differs to these connection mechanisms in the sense that we do not prescribe a particular mechanism in how influence is spread, instead, we ask the RL agent to intelligently find the best strategy/policy to achieve influence spread by saying the right words and causing connections to be formed or removed based on consensus. This means it always has to adapt its output based on how the influence is being spread currently. Perhaps statistically this eventually leads to a diffusion-like mechanism, but this is beyond the scope of our scientific goals.

Research Focus	Data / Environment	Influence Mechanism & Agent Design
Debate & Arguments (Betz 2022)	Verbal Simulation	Symbolic reason balance model
Spatial influence (Jaques et al. 2019)	Maze world	RL tasks
Opinion dynamics (Siahkali et al. 2024)	SBC Model	Deep RL
Generative diffusion (Ghaffarzadegan et al. 2024)	Simulation	Diffusion with LLM
Explicit dynamic graphs (Jin & Guo 2022)	OSN data sets	Graph-ODEs & Neural Nets (Zou et al. 2024)
Language informed tasks (Panagopoulos et al. 2024)	Simulation	LLM driven Q-values in RL
Real-world experience (Park et al. 2023)	Simulation world	Memory-stream & Reflection driven LLM
Our proposed work	OSN topic data sets	RL with LLM output

Table 1: Review of Current Work and Positioning of Our Work. Our key contributions are: (i) topic specific real-data, and (ii) no assumed influence mechanism - we utilize RL-driven influence strategies.

Agent design using reinforcement learning (RL)

2.4 As previously discussed, in RL agent design, the target agent learn optimal strategies based on the interaction with social environment. Originating from animal learning in psychology, RL has the capacity to mimic the human-like action selection. In 1989, Watkins & Dayan (1992) significantly advanced the field by integrating the theory of optimal control with temporal-difference (TD) learning, leading to the development of Q-learning. To address the curse of dimensionality in Q-learning, Mnih et al. (2015) combined deep learning with RL to create the deep Q-network (DQN), which notably surpassed professional human players in a range of 49 classic Atari games. We know from evidence that social conformity and influence is closely associated with RL and hence our design is motivated by evidence from Klucharev et al. (2009), and used in advertising to increase social influence modeled via Stochastic Bounded Confidence Model (SBCM) (Siahkali et al. 2024).

2.5 Recent advances have extended the application of RL in various domains in natural sciences, social sciences, and engineering (Silver et al. 2016; Isbell et al. 2001; Luo et al. 2017). In the work of Sert et al. (2020), authors show that the combination of RL and ABM can create an artificial environment for policy makers to observe potential and existing behaviors associated with the rules of interactions and rewards. These RL agents often involve using graph embeddings as a way to articulate the importance of network centrality in social influence mechanisms (Wang et al. 2021), such as diffusion and consensus through synchronization (Zou et al. 2024). A challenge in current literature is that often the opinion or influence is mathematically modeled in a general way. For example, Jin & Guo (2022), Jin et al. (2025) projected online social network (OSN) data through NLP dimension down-scaling onto ordinary differential equation (ODE) dynamic social graphs. These approaches (Shepherd & Goldsmith 2020) lack a more naturalistic interface that is compatible with how humans interact and as such also lack topic-sensitivity. RL agents have also been actively applied to social robotics to exert influence (Ndousse et al. 2021), but the interplay with LLMs to enable more naturalistic behaviour and human interaction has been a growing recent research topic. There are also spatial applications of RL driven influence as shown in (Jaques et al. 2019), where robots use their spatial navigation and task solving ability to create influence on others. The integration of LLM into RL has also recently advanced, for example using LLM knowledge to inform the attention space (e.g., Q-values) of RL agents (Panagopoulos et al. 2024).

Large language model

2.6 LLM such as GPT has been tested for multiple natural language processing tasks and proved to be effective without gradient updates or fine-tuning (Brown et al. 2020). Since GPT-3 was trained on large web corpora which include social media behavior, GPT-3 can predict user's future behaviours, responses, or action plans with some accuracy. GPT was able to give human-like response to replace the expensive and time-consuming user studies (Sekulić et al. 2022). Although the generated synthetic data was proved to exhibit less language variability, it may lead to better classification results when few data is available and resources are limited (Meyer et al. 2022). Recent study has use Large Language Models to simulate the online social interactions, leading to improved dialogue and outcomes through debate (Du et al. 2024). LLMs are used to predict the popularity on the online social media (Park et al. 2022), the distribution of votes for elections, the sentiment of news, and the voting of justices (Hamilton 2023). In the work of Park et al. (2022), authors presented a social simulacra tool, SimReddit, which is a prototyping technique generating social behaviours using large language models and matching to complexity analysis (Lu et al. 2024). With the description of a community's design, the social simulacra produce appropriate simulated behavior. Game theoretic frameworks have also been explored using LLMs (Lore & Heydari 2024).

Simulated Social Environment

3.1 One challenge faced by researchers is the ethics of real-world experimentation and the lack of social influence data. The purpose of the simulated environment is to better understand how online social members influence each other through dialogue and follow/unfollow mechanisms. Our environment contains agents that generate posts, form opinions on topics, and socially follow each other based on the outcome of discussions. This simulation allows researchers to observe the evolution of how opinions form and how influence leaders emerge. In the following section, we detail how an agent can employ Reinforcement Learning (RL) to optimise its interaction with the community with the purpose of maximizing its influence and following. We then detail the results and findings of our computation experiments.

Environment

3.2 In this section, we introduce the development of a simulated environment that replicates the dynamics of group discussions on online social media platforms. Discussions in this environment revolve around the topic, defined by posts that represent divergent viewpoints within the conversation.

3.3 The simulated environment system graph is shown in Figure 1a. We will explain each part of the system step by step.

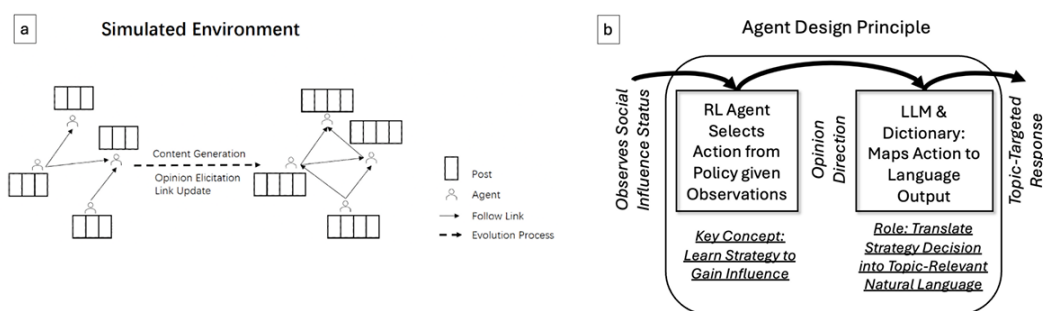


Figure 1: (a) Simulated Environment Social Network: Participants within this environment, referred to as 'agents', are interconnected through the 'follow' relationship between each other. Agents interact through 'Posts' which is a small natural language message, formatted like a Tweet. Social influence evolution process shows the development of the whole network from t to $t + 1$. The process can be segmented into three part: Post Generation, Opinion Elicitation and Link (follow) Update. This process is how we track and measure influence. (b) Agent Design Principle: The agent is comprised of 2 building blocks - (i) the RL model that formulates how best to strategically gain influence in the long term, and (ii) the LLM that maps the strategy decisions from RL to natural language tweets for the topic specified.

- 3.4 Within this simulated setting, a post is a small natural language message, formatting like Tweet. These posts are contributing to the whole discussion, each either supporting or opposing the central topic.
- 3.5 For instance, in our experiment, we initiate discussions using the topic: 'Society with no gender'. Posts such as 'This incorrectly supposes that some current biological sex traits aren't functional. I think a society without gender is a really bad idea.' are categorized as 'con' post to the topic. Conversely, posts like 'The absence of gender constructions would enable people to enjoy a better overall quality of life.' support the topic, as a 'pro' post.
- 3.6 Participants within this environment, referred to as 'agents', are interconnected through the 'follow' relationship between each other. Follow relationship is unidirectional. Each agent generates a post $POST_i^t$ at time t , influenced by their existing follow links and previous posts. Consequently, if agent i follows agent j at time t , then $POST_j^t$ potentially impacts $POST_i^{t+1}$.
- 3.7 At any given time t , $POST_i^t$ fully determines the agent i 's opinion $OPIN_i^t$. The network of follow links between agents is then dynamically updated in response to these evolving $OPIN^{t+1}$.
- 3.8 The process can be segmented into three part: Post Generation, Opinion Elicitation and Link Update, as outlined in the following Algorithm 1.

Algorithm 1 Simulated environment algorithm.

```

for  $t \in [1 \dots t_{\max}]$  do
  for  $i \in \text{AGENTS}$  do
    locate the follow link of agent  $i$ 
    generate the post of agent  $i$  ( $\rightarrow POST_i^t$ );
    elicit the opinion of agent  $i$  ( $\rightarrow OPIN_i^t$ );
  end for
  update the follow link of AGENTS;
end for

```

Post generation

- 3.9 This section explores the mechanics of LLMs in content generation technology. These models predict the next word in a sequence with a probabilistic approach given the prompt input.
- 3.10 The objective of the post generation stage is to generate the possible post of agent at the given time. This process incorporates the agent's previous posts and those from their followees. The post is generated by the LLM based on instructions received from RL model action space (see Figure 1b). As such, the LLM is not involved in action selection of the RL, rather it is a plug-in that translates opinion action space into natural language for dialogue
- 3.11 The generated post must be concise a short sentence formatting like Tweet. Therefore we use the natural language description to describe the follow relationship, choose the format of web API including a designated stop sign as ('}') to ensure the output aligns with the expected structure.
- 3.12 Consider the following example: if agent i follows agent j , the prompt for generating $POST_i^{t+1}$ would be:


```

agenti posted tweets about 'a society without gender': POSTit
agenti saw the following tweets about 'a society without gender' on homepage{'data':{'user': agentj, 'text': POSTjt }}. "
agenti shared the following tweet on this topic: {'data: {'user': agenti, 'text': "" }}

```
- 3.13 The first paragraph of the prompt describe the topic and refer the previous posts from agent i . The second paragraph then suggests agent i follows agent j , so the post of agent j will potentially impact the next post of agent i . If agent i follows multiple agents in the group, the prompt will include more content that agent i can read from the homepage. The third paragraph ensure agent i still discuss the same topic and generation of tweet-like post. The name of agents, allowing for the observation of interaction between them, such as agreements or disagreements on the topic. An example of a generated post, demonstrating such an interaction, is as follow:

"I agree with both Grace Lee and Chloe Kim. Gender is an important aspect of our identity and should not be disregarded. However, everyone should also have the freedom to express themselves and live authentically. Instead of erasing gender, we should focus on promoting equality and challenging harmful gender stereotypes. Let's create a society where individuals are free to express their gender identity without fear of discrimination or prejudice. #GenderEquality #BeYourself"

- 3.14** This response reflects incorporating with other agents and presenting a balanced perspective. It demonstrates that the generation of posts is not only topic-relevant but also reflective of the influences in a social media environment.
- 3.15** Moving to text generation process, key stages include 'encoding', where the model assimilates and interprets the input text, identifying contextual patterns, and 'decoding', where it generates new text from this analysis.
- 3.16** Crucial to the flexibility of these models in the decoding phase are key parameters such as 'temperature' and 'top_p'. Temperature controls the randomness in the prediction, influencing creativity and coherence in the generated text. 'top_p', controls the model to focus on a subset of the most probable next words, ensuring relevance and precision in the output. These two parameters balance the generated content between novelty and fidelity to the input context.
- 3.17** Table 2 shows the our setting for narrow case and creative case. Low temperature and low top-p lead to conservative, and narrow minded agents who're sticking with the most obvious options when generating a post. The creative setting characterize agents who are much willing to take surprising turns.

	temperature	top-p
Narrow	0.1	0.5
Creative	1.4	0.95

Table 2: GPT Decoding Setting.

Opinion elicitation

- 3.18** The opinion elicitation process gives the probability that the agent expects a pro-claim rather than a con-claim given the perspective using perplexity calculation. The method we used is introduced in the work of Betz Betz (2022), where we follow its steps but with a different purpose. Betz aimed to discover whether introducing new arguments will disrupt the dynamics of collective deliberation, whereas our paper aims to improve the number of followers/social influence by introducing a RL-LLM agentic intelligence to our agent.
- 3.19** Perplexity effectively measures how well a probabilistic model predicts a sample, serving as an indicator of the model's predictive capacity across the entire set of tokens within a corpus. In the context of natural language processing, perplexity is defined as the exponential of the average negative log-likelihood of a given sequence of tokens. Consider a tokenized sequence represented as $S = (w_0, w_1, \dots, w_m)$. The perplexity of S denoted as $PPL(S)$, is computed as:

$$PPL(S) = \exp \left(-\frac{1}{m} \sum_{k=0}^m \log p_{\theta}(w_k | w_{<k}) \right) \quad (1)$$

- 3.20** In Equation (1), $\log p_{\theta}(w_k | w_{<k})$ represents the logarithm of the likelihood of the k th token in the sequence, given all preceding tokens (denoted as $w_{<k}$) as per the probabilistic model in use.
- 3.21** When GPT processes a sentence, it calculates the logarithm of the probability of each word given the previous words in the sentence, which is $\log p_{\theta}(w_k | w_{<k})$ in the equation.

$$PPL_{CON}(POST_t^i) = PPL(S_{CON} | POST_t^i) \quad (2)$$

$$PPL_{PRO}(POST_t^i) = PPL(S_{PRO} | POST_t^i) \quad (3)$$

- 3.22** Equation (2) shows how to compute $PPL_{CON}(POST_t^i)$. Given a predefined sentence S_{CON} like "society without gender is a really bad idea", the conditional perplexity shows the inverse probability of the output of the language model. Similarly, $PPL_{PRO}(POST_t^i)$ is computed using the predefined sentence "Society without gender is a really good idea". The low perplexity result suggests that the language model predicts this particular sentence with a high level of certainty.

$$OPIN_t^i = \frac{PPL_{CON}(POST_t^i)}{PPL_{PRO}(POST_t^i) + PPL_{CON}(POST_t^i)} \quad (4)$$

3.23 Based on the conditional perplexity scores for pro and con sentences, which indicate the confidence of the language model on each side of perspective, we compute the overall opinion score. In Equation (4) we give the definition of $OPIN_i^t$. The outcome is a float number in range $[0, 1]$ which represent the polarity where 0 represents con and 1 represent pro. We will then categorise the number into five polarity categories: 'strong con', 'con', 'neutral', 'pro' and 'strong pro'.

Link update

3.24 To establish the 'Following' relationship between users, we adopt the premise that individuals inclined to follow users who express similar opinions, as suggested by Zhou et al. (2009). Once we have the categorised polarity of each tweet, we then hypothesise that individuals with the same polarity category hold similar opinions.

3.25 Therefore, we explore two link methods: a Follow Dynamics and a Follow-Unfollow Dynamics. In both cases, if two users consistently align in the same opinion category over a few consecutive steps, they are considered likely to follow each other, with a follow probability. In Follow-Unfollow Dynamics, the exist follow link may also break with an unfollow probability, if the opinions of two people differ during the consecutive steps. Details of follow link update and unfollow link update are given in Algorithm 2 and Algorithm 3.

Algorithm 2 Follow link update algorithm.

```

for each agent  $i$  do
  for each agent  $k \neq i$  not already followed by  $i$  do
    for  $t \in [1, \text{num\_same}]$  do
      if  $OPIN_i^t$  and  $OPIN_k^t$  in the same opinion category then
        continue
      else
        match  $\leftarrow$  false
        break
      end if
    end for
    if loop completed without break then
      if random number  $<$  follow link rate threshold then
        match  $\leftarrow$  true
        Add link from  $i$  to  $k$  in new follow dict
      end if
    end if
  end for
end for

```

Algorithm 3 Unfollow link update algorithm.

```

for each agent  $i$  do
  for each agent  $k \neq i$  already followed by  $i$  do
    for  $t \in [1, \text{num\_same}]$  do
      if  $OPIN_i^t$  and  $OPIN_k^t$  not in the same opinion category then
        if random number  $<$  unfollow link rate threshold then
          match  $\leftarrow$  false
          Delete link from  $i$  to  $k$  in new follow dict
          break
        end if
      end if
    end for
  end for
end for

```

● Agent Design: Reinforcement Learning

4.1 In the previous part, we have introduced the whole simulated environment. In this part, we detail the reinforcement learning setting employed for the model.

Model design

4.2 We set a target agent in this group network without any initial followers. Then the target agent can actively choose current opinion state and post related posts. The objective is try to let the target agent get more followers within the determined time step.

4.3 The reinforcement learning (RL) model utilized in this framework is based on the q-learning algorithm, a widely used method in machine learning for environments characterized by uncertainty. This technique is used for developing optimal strategies through learning from interactions within such environments.

4.4 The RL Loop structure is shown in Figure 2. The target agent interacts with the previous introduced simulated environment. At each time step, the agent observes the state of the environment, selects an action, receives a reward from the environment, leading to a new environmental state.

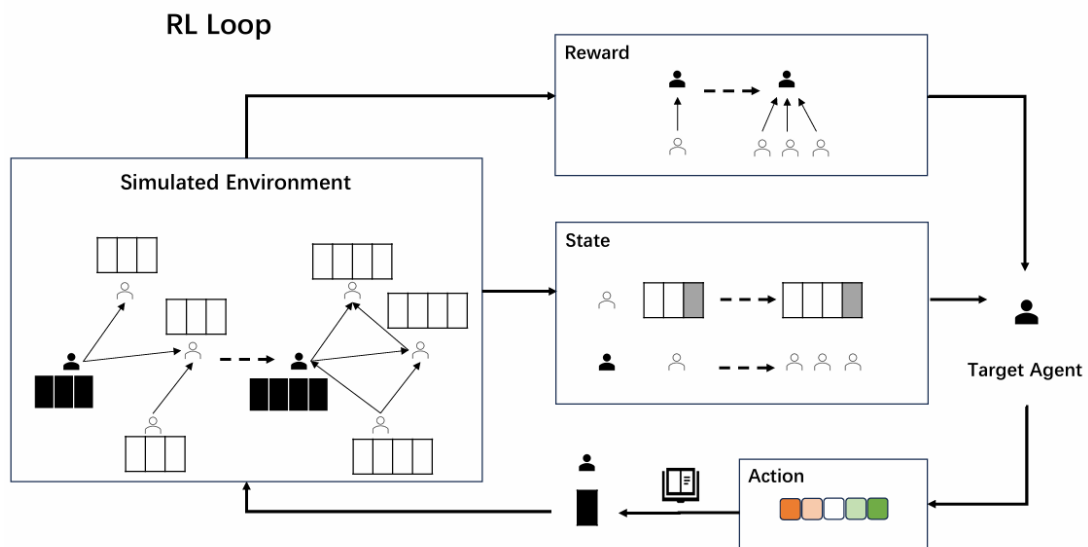


Figure 2: RL Loop. Reward: change of number of followers at each time step. State: 1) the current opinion states of the initial accounts followed by the target agent, 2) the number of followers the target agent currently has. Action: pre-defined five opinion polarity categories.

Reward

4.5 The reward mechanism for this model is the change of number of followers at each time step, encouraging strategies that enhance follower acquisition.

State

4.6 To determine what a target agent can observe, we assume that the agent has visibility over the accounts it follows, as well as its own state. The observed state is a matrix which comprises two elements: 1) the current opinion states of the accounts followed by the target agent, 2) the number of followers the target agent currently has.

Action

- 4.7** At the action stage, target agents participate in the discussion by selecting a sentence to express their opinion. The available action space consists of five predefined opinion polarity categories: strong con, con, neutral, pro, and strong pro. The LLM is not involved in action selection of the RL as shown in Figure 1b, rather it is a plug-in that translates opinion action space into natural language for dialogue.
- 4.8** In our experiment, we use the actual posts to generate an action dictionary. The dictionary maps actual posts to opinion categories. After selecting an opinion category based on the RL algorithm, the target agent randomly chooses a tweet from the corresponding category.

Reinforcement learning algorithm

- 4.9** Referencing Algorithm 4, the process begins with the initialization of Q-values for all state-action pairs. In each episode, the state is initialized, and the agent embarks on a series of time steps. During each time step, the target agent selects an action from its current state, based on a policy derived from the Q-values. Alongside each agent in the system generates posts and forms opinions. The algorithm then updates the follow links among agents and processes the resulting reward (followers) and new state. This leads to an update in the Q-value for the chosen action, and the agent's state is updated accordingly. The Q-value update is represented by the formula:

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[r + \gamma \max_{a'} Q(s', a') - Q(s, a) \right] \quad (5)$$

- 4.10** In this formula, α represents the learning rate, γ the discount factor, r the reward (followers), and $Q(s', a')$ the maximum estimated future reward. This update rule is key to the learning process, as it adjusts the Q-values based on the reward received and the potential future rewards, leading to more informed and strategic decision-making by the agent in subsequent steps. The details of the RL design and parameter tuning can be found in Table 3. We use Gender case and Drug case, set number of episodes 500, max of steps per episode as 5. Other opinion experiment specific parameters are shown in Appendix-D.

Algorithm 4 Reinforcement Learning Algorithm.

```
Initialize Q-values  $Q(s, a)$  arbitrarily for all state-action pairs  $(s, a)$ .
for each episode do
  initialize state  $s$ 
  for  $t \in [1 \dots t_{\max}]$  do
    target agent choose action  $a$  from state  $s$  using policy derived from Q.
    for  $i \in \text{AGENTS}$  and  $i$  is not target agent do
      generate the post of agent  $i$  ( $\rightarrow \text{POST}_i^t$ );
      elicit the opinion of agent  $i$  ( $\rightarrow \text{OPIN}_i^t$ );
    end for
    update the follow link of AGENTS;
    observe reward  $r$ , and new state  $s'$ ;
    Update Q-value for  $(s, a)$ ;
    Update state  $s \leftarrow s'$ ;
  end for
end for
```

Agentic RL Design Parameter / Function	Design Approach / Value
RL type	Q-learning
reward function	the number of followers
observation states	opinion states of other agents followed, number of followers, self-observation
action space	5 levels of opinions
data case studies	(1) gender debate, (2) drug-use debate
learning rate	0.01
discount rate	0.99
epsilon	0.005
exploration rate	0.01
max. steps per episode	5
no. episodes	500

Table 3: Q-learning RL Design and Training Parameters.

Experiment

Simulated environment initialisation

- 5.1** The initiation of our simulated environment begins with the creation of initial posts, for which we use the dataset released by Betz (2022). The dataset is crawled from debating platform kialo.com, comprises various posts structured around specific topics. For our simulation, we select posts related to the topic 'Gender' and topic 'Drug'. The dataset includes 486 posts under 'Gender' topic and 660 posts under 'Drug' topic, with each posts labeled with one type 'pro' or 'con'. The length of each post is less than 70 words. For example, under 'Gender' topic, a post such as 'The absence of gender constructions would enable people to enjoy a better overall quality of life.' is labeled as type 'pro'.
- 5.2** Employing the opinion elicitation process, we analyze the datasets to determine the polarity distribution. As shown in Figure 3, the distributions of polarity value in both cases cluster around the middle range. This distribution is indicative of the varying degrees of support or opposition to the topic within this dataset.

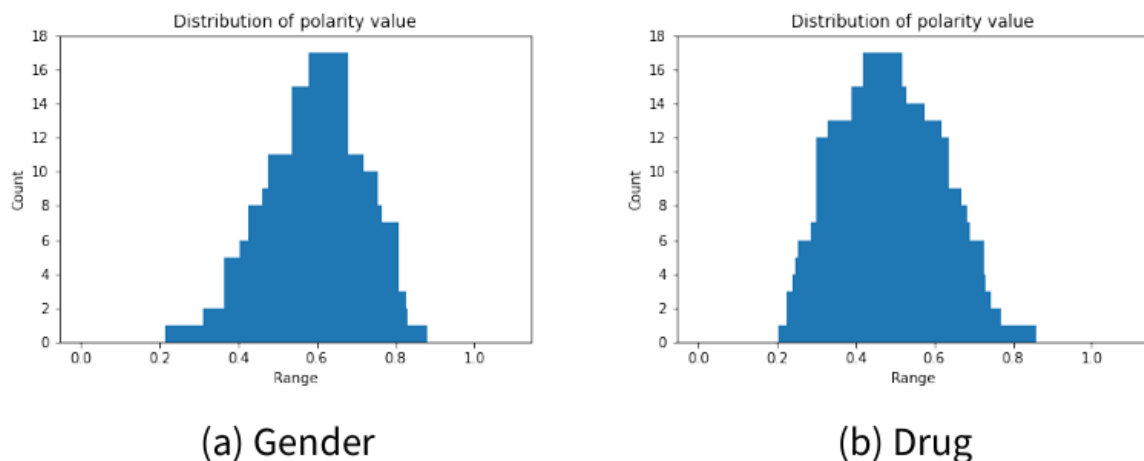


Figure 3: Distribution of polarity value.

- 5.3** Based on the polarity distributions, we categorize the posts into distinct opinion groups while ensuring an even distribution across these categories. The specific range settings are detailed in Appendix D, and the categorization results are presented in Table 5 and Table 6.

Simulated environment analysis

- 5.4** To examine the opinion evolution process within our simulated environment, we run two 50-step simulations using both Narrow and Creative agent setting using Gender topic. These simulations aim to demonstrate how different agent settings can influence the dynamics of the discussion.
- 5.5** Figure 4 illustrates the dynamics of opinion evolution under the influence of two simulation configurations: the Narrow setting and the Creative setting. Table 2 provides the specific Temperature and Top-p settings used for the opinion dynamics simulation.

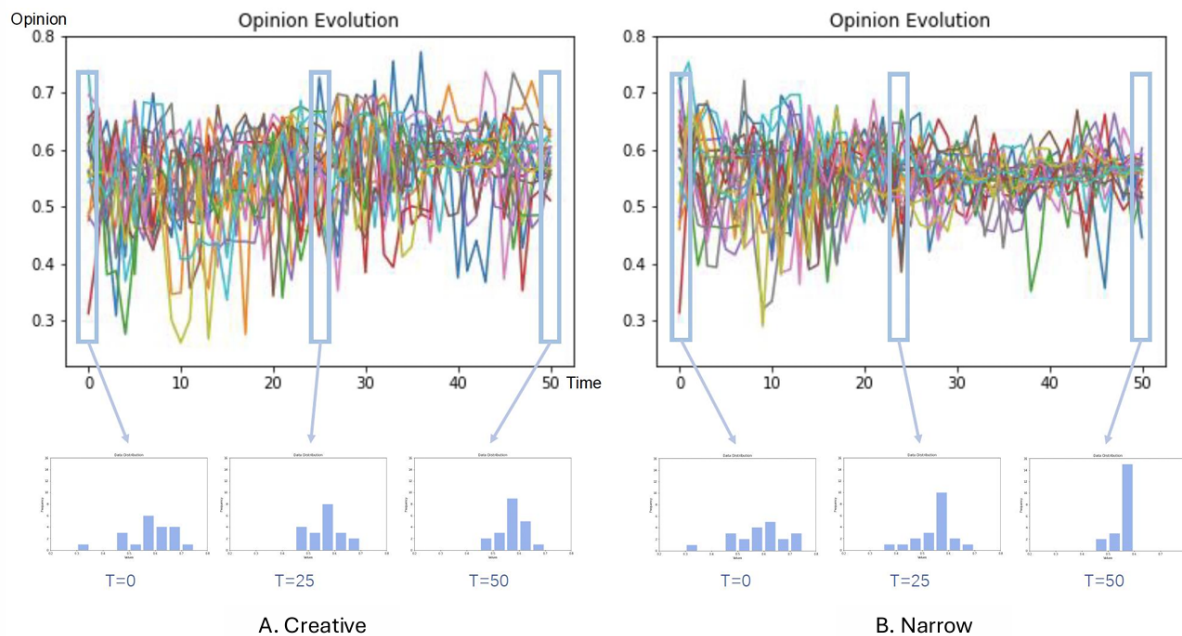


Figure 4: Opinion Evolution Visualization on Narrow setting and Creative setting.

- 5.6** Opinion evolution visualizations show the trajectory of opinion polarity score for 20 agents over 50 steps on Narrow setting and Creative Setting. Initially, the agents' opinions range from 0.3 to 0.8. In the Narrow setting, we observe that opinion density increases over time, showing a tendency toward opinion convergence. Conversely, in the Creative setting, the opinion density remains more dispersed and chaotic throughout the entire time scale, with less apparent convergence.
- 5.7** To provide a more detailed view of opinion dynamics, opinion distributions at Time 0, 25, and 50 are presented. In the Creative setting, despite the initial high degree of opinion diversity, after 25 steps, some grouping starts to occur, but full convergence is not reached even by step 50. The discussion remains dynamic, with polarity scores fluctuating in the noisy range between 0.4 and 0.7. On the other hand, in the Narrow setting, while there is significant opinion diversity in the first 25 steps, this diversity gradually decreases, and by step 50, opinions cluster in the 0.45 to 0.6 range, indicating a clearer sign of convergence.
- 5.8** In both scenarios, the distribution of the final polarity scores shows a slight skew towards higher values across the time steps.
- 5.9** Figure 5 plots the standard deviation of opinion over time in both Narrow and Creative settings. The graph shows that in the Narrow setting, the opinions form a more centralized distribution by step 50, indicating stronger clustering and a more definitive convergence of opinions, as compared to the Creative setting, which remains more dispersed throughout the simulation.

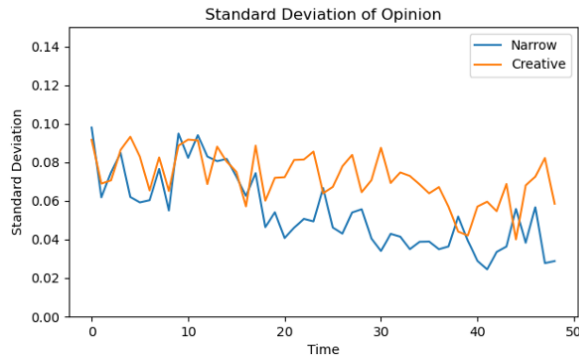


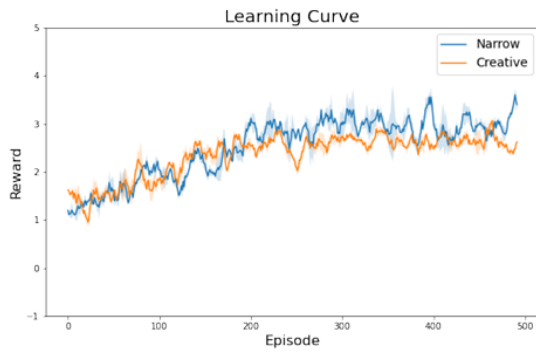
Figure 5: Standard Deviation of Opinion on Narrow setting and Creative setting.

Reinforcement learning setting

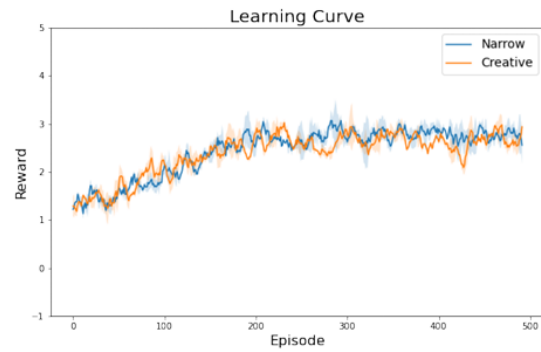
- 5.10** In this section, we explore the application of RL within our simulated environment. The primary research question is: *Can RL algorithms identify an optimal strategy that enables the target agent to maximize its follower count within a predetermined number of steps?*
- 5.11** By testing different settings, we found that limiting the action space of the RL agent and incorporating self-observation are key factors for stable RL result. We found that actions when restricted by using only statements in the dictionary, and not the LLM, significantly improved convergence in training due to the narrow topic of our experiments (e.g., gender rights).
- 5.12** By restricting the action space, we reduce the complexity of the decision-making process (e.g., number of actions in a policy), enabling the RL agent to focus on a more manageable set of actions. This helps prevent the model from exploring either irrelevant or suboptimal actions within a policy that could introduce noise and instability into the learning process. Additionally, a smaller, well-defined action space allows for faster convergence, as the agent can more efficiently learn the consequences of its actions within the environment.
- 5.13** Incorporating self-observation further enhances stability by allowing the agent to continuously assess its own state and progress. This self-awareness helps the agent make more informed decisions, as it can adjust its behavior based on how past actions have affected its follower count and other relevant metrics. Self-observation also promotes consistency, as the agent is better equipped to identify patterns in the environment and avoid erratic or random behaviors. Together, limiting the action space and incorporating self-observation contribute to generating more stable and reliable outcomes, ensuring that the RL agent can effectively optimize its strategy within the defined constraints of the simulation.
- 5.14** To investigate stable RL results, we conducted tests across eight scenarios.
1. **Language Model Settings:**
 - We employed two types of settings for the language models of agents, as delineated in Table 2: *Narrow* and *Creative*. These settings impact the nature of posts generated by the agents.
 2. **Link Update Methods:**
 - *Follow Dynamics*: In this case, the follow rate is 0.8, and the unfollow rate is 0. This setting allows us to observe the growth of the target agent’s network under a scenario where new follow links are relatively likely to form.
 - *Follow-Unfollow Dynamics*: In this more dynamic setting, the follow rate is still 0.8, and the unfollow rate is set as 0.5. This case allowing existing follow links to break, thus adding a layer of complexity to the network dynamics.
 3. **Observability Settings:**
 - *Part-Observable Case*: In this scenario, the target agent initially follows only one other agent.
 - *Full-Observable Case*: Contrasting the Part-Observable setting, in this case, the target agent begins by following every agent in the network.
- 5.15** We use Gender case and Drug case, set number of episodes 500, max of steps per episode as 5. Other parameters are shown in Appendix-Parameters.

Reinforcement learning results

5.16 This subsection presents the outcomes of our RL experiments conducted across the eight scenarios. Figure 6 illustrates the results for the Gender scenario and Figure 7 shows the results for the Drug scenario. In both figures, the colored lines represent the average reward (followers) of five repeated experiments, and the shaded areas indicate the variance in the results. The followers gained (reward) rises from 1 to 2 or 3 in different cases in a total agent population of 20. That is to say, we see a gain of 10% of the total available follower population.



(a) Follow Dynamics, Part-Observable



(b) Follow Dynamics, Full-Observable



(c) Follow-Unfollow Dynamics, Part-Observable



(d) Follow-Unfollow Dynamics, Full-Observable

Figure 6: Learning Curves of RL in Gender case, with different language model settings, link update methods, and observability settings.

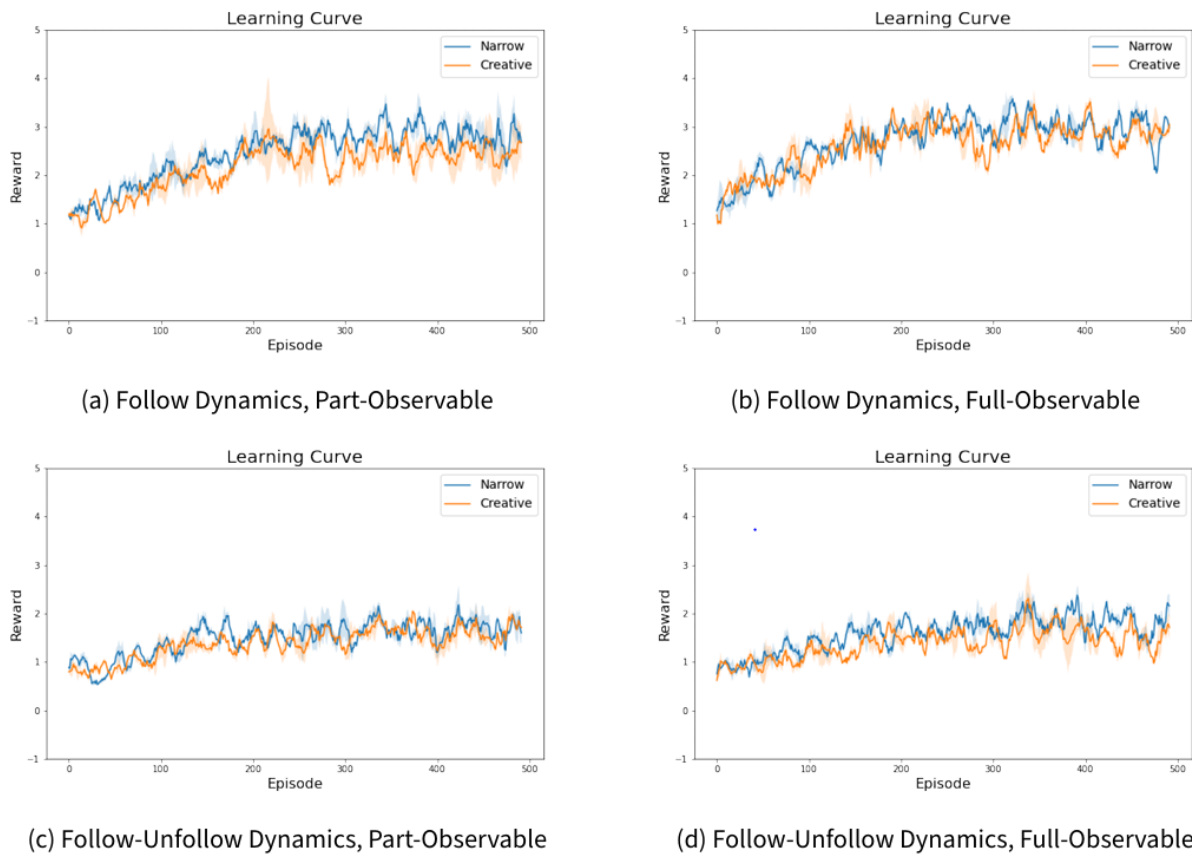


Figure 7: Learning Curves of RL in Drug case, with different language model settings, link update methods, and observability settings.

- 5.17** Regarding a baseline, if there was no intelligence from RL and we just used initial conditions, it would be a statistical flat line with no followers gained over time.
- 5.18** In all scenarios, we observe convergence in the reward learning curves (followers), indicating that the RL algorithm successfully identified optimal solutions within the given settings. Both the Gender and Drug cases exhibit similar learning curve patterns under different conditions, highlighting consistent performance across varying contexts.
- 5.19** One notable observation is that all Follow-Unfollow Dynamics cases exhibited greater variance in their learning curves compared to the Follow Dynamics counterparts under same setting. This implies that more complex scenarios pose greater challenges in reaching optimal results.
- 5.20** Cases with Full-Observable settings demonstrates more stable learning curve than the part-observable ones. This suggests that target agent with complete network visibility are likely to learn strategies more effectively and with greater stability than those with limited visibility.
- 5.21** The Narrow and Creative settings, when applied to the Partially-Observable cases, presented distinct learning curves. The Narrow setting consistently outperformed the Creative setting, as evidenced by higher reward values. However, in the Full-Observable scenarios, the learning curves for both Narrow and Creative settings align closely. A possible explanation is that, in Part-Observable scenarios, the unpredictable nature of agents in the Creative setting, leading to lower performance compared to the Narrow setting. Conversely, in Full-Observable scenarios, the agents are better able to adapt to and anticipate the Creative setting's variable patterns.
- 5.22** In summary, these results highlight the impact of observability and complexity in reinforcement learning environments, with Full-Observable settings and simpler dynamics facilitating more stable and efficient learning outcomes. The Narrow setting's predictable nature resulted in higher performance than the unpredictable Creative setting. However, when observability was not a constraint, as in the Full-Observable scenarios, both settings showed similar performance. This suggests that with full visibility, agents can effectively adapt to even unpredictable dynamics.

● Conclusions and Future Work

- 6.1 Understanding the dynamics of public opinion evolution on online social platforms is crucial for understanding influence mechanisms and the provenance of information. Traditional influence analysis is typically divided into qualitative assessments of personal attributes (e.g., psychology of influence) and quantitative evaluations of influence power mechanisms (e.g., social network analysis).
- 6.2 Many of the analysis approaches assume how influence is spread (e.g., diffusion, Ghaffarzadegan et al. 2024; opinion ODE Jin & Guo 2022, or stochastic bounded confidence models, Siahkali et al. 2024), but these are not topic-specific and lack dynamic adaptation at the individual scale. Our approach uses reinforcement learning (RL) to strategically decide what actions to take in conversations and an LLM to translate the actions into natural language. This enables us to better modulate our conversations to be natural and topic-specific, whilst maximizing the individual agent's long-term influence upon peers. Whilst there is emerging work on integrating LLMs and RL agents (Panagopoulos et al. 2024) and large-scale simulations of agent behaviour (Park et al. 2023), we believe at time of writing this is an important contribution.
- 6.3 In terms of academic contribution, our work also enable researchers to ethically test new agents and create fine-scale experimentation on opinion dynamics for diverse topics.
- 6.4 Our current findings reveal the following: (1) RL can achieve a gain of 10% of the total available follower population, and (2) convergence can be better achieved by limiting the action space and incorporating self-observation. This demonstrates the simulation framework's capacity to create agents that can adapt to complex and unpredictable social dynamics. The potential application would be providing valuable insights for digital marketing strategists and policymakers in the dynamics of online opinion formation and influence. However, we acknowledge that one potential negative social impacts could be the misuse of the simulated environment in manipulating public opinion. If used unethically, the technology may craft highly persuasive and targeted misinformation campaigns on social media platforms. Conversely, this technology can also be instrumental in detecting malicious activities in social networks.
- 6.5 One limitation of our research is the inherent unpredictability of actual social media discussions. Social dynamics are complex and unpredictable, and our rule-based simulated environment cannot fully replicate. Although the pre-trained RL algorithm demonstrates promising performs in the simulated environment, it presents one version of the actual possible outcome. Despite this, our research provides valuable insights into the mechanics of social interaction and opinion dynamics in online groups.
- 6.6 Our future work would be directed towards enhancing the realism and variability of our simulated social media environment. This may involve incorporating more sophisticated interaction rules to better mirror the real-world social media dynamics. To achieve this, first we plan to discover how different character profiles of agents perform and interact within the network. This exploration aims to provide a deeper understanding of the varied behaviors and their impacts on the overall dynamics of social media interactions. We can also improve the follower mechanism, where we can create more complex multi-topic link updates, where agreement on one topic, might lead to a following that is across all similar topics – reflecting closer to the way social connections form. This can be expanded to cover similar or shared geographies, hobbies and, lifestyles. Currently, we are limited by our real data sets on topics. We are in particular interested in any ability to model social acceptance of future emerging technologies (e.g., 6G), where there are no real data sets, but understanding social opinion is critical. We do note that there are emerging synthetic world simulations (e.g., academic Park et al. (2023) and commercial Project OdySSEy) to emulate a realistic life and social interactions.

● Code & Data Availability Statement

The source code (Python) for running the simulations is published at: <https://github.com/AlminaJin/InfluentialAgent>. All data is available from public data bases.

● Acknowledgments

The work is supported by: (1) "Networked Social Influence and Acceptance in a New Age of Crises" funded by USAF OFSR (FA8655-20-1-7031); and (2) "Communications Hub For Empowering Distributed Cloud Computing Applications And Research" funded by EPSRC/DSIT (EP/X040518/1) (EP/Y037421/1).

● Appendix A: GPT API Setting

Table 4 outlines the parameters set for the GPT API, distinguishing between 'creative' case and 'narrow' case configurations. These parameters are used in LLMs to tune how much creativity they add to the natural language output. The role of LLM is only to map the RL opinion decision to a natural language format. Therefore, this creativity is only in the mapping stage, not in the decision stage. Most literature (Park et al. 2022) use these parameters similarly to map the boundaries of the LLM performance. Hopefully this explains that this does not affect how opinion influence decisions are made, but only the text format in which it manifests. Within these settings, higher temperature and top-p values are indicative of a setup geared towards fostering creativity. For tasks related to post generation and opinion elicitation, we employ the pre-trained GPT-3.5 model developed by OpenAI (OpenAI 2022).

	Creative	Narrow
max_tokens	2000	2000
temperature	1.4	0.1
top_p	0.95	0.5
n	1	1
stop	"}"	"}"
model	"gpt-3.5-turbo"	"gpt-3.5-turbo"

Table 4: GPT Creative Parameters.

Appendix B: Opinion Example

In this section, we will explore illustrative examples of 'strong pro' and 'strong con' concerning the concept of a 'society with no gender.' These examples are meant to elucidate the polarized viewpoints on this topic.

A 'strong pro' argument is presented as follows:

"Any violent behavior targeted at the traits now associated with gender would be illogical, because that would assume that gender stereotypes would persist without gender. Such an assumption would almost certainly also include the academically false assumption that gender is biological."

Conversely, a 'strong con' perspective is showed in the following statement:

"It won't disappear. The idea of Sexist Violence in the eyes of society might, but that's only because what we call Gender-based violence is just that... Violence. What you would see in a drop in Gender based violence, there would be a spike in generalized Violence to take its place. Its only taking a tag of something and putting it into a generalized pool of everything else. It helps no one and Dis-benefits everybody involved. So a society without gender is a really bad idea."

Appendix C: Opinion Category Setting

Based on the polarity distributions, we categorize the posts into distinct opinion groups while ensuring an even distribution across these categories. The specific range settings and the categorization results are presented below.

Category	Polarity score	Number of posts
Strong pro	1 - 0.64	83
Pro	0.58-0.64	89
Neutral	0.51-0.58	128
Con	0.45-0.51	93
Strong con	0 - 0.45	93

Table 5: Gender opinion category setting.

Category	Polarity score	Number of posts
Strong pro	1 – 0.55	101
Pro	0.48–0.55	128
Neutral	0.42–0.48	145
Con	0.37–0.42	102
Strong con	0 – 0.37	184

Table 6: Drug opinion category setting.

Appendix D: Model Documentation

The source code (Python) for running the simulations is published at: <https://github.com/AlminaJin/InfluentialAgent>

References

- Betz, G. (2022). Natural-language multi-agent simulations of argumentative opinion dynamics. *Journal of Artificial Societies and Social Simulation*, 25(1), 2
- Briguglio, L., Nesse, P.-J., Di Giglio, A., Occhipinti, C., Durkin, P. & Markopoulos, I. (2021). Business value and social acceptance for the validation of 5G technology. 2021 IEEE International Mediterranean Conference on Communications and Networking (MeditCom)
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I. & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901
- Du, Y., Li, S., Torralba, A., Tenenbaum, J. B. & Mordatch, I. (2024). Improving factuality and reasoning in language models through multiagent debate. *Proceedings of the 41st International Conference on Machine Learning*
- Ghaffarzadegan, N., Majumdar, A., Williams, R. & Hosseinichimeh, N. (2024). Generative agent-based modeling: An introduction and tutorial. *System Dynamics Review*, 40(1)
- Hamilton, S. (2023). Blind judgement: Agent-based supreme court modelling with GPT. *arXiv:2301.05327*
- Isbell, C., Shelton, C. R., Kearns, M., Singh, S. & Stone, P. (2001). A social reinforcement learning agent. Available at: <https://www.cis.upenn.edu/~mkearns/papers/cobotRL.pdf>
- Jackson, J. C., Rand, D., Lewis, K., Norton, M. I. & Gray, K. (2017). Agent-based modeling: A guide for social psychologists. *Social Psychological and Personality Science*, 8(4), 387–395
- Jaques, N., Lazaridou, A., Hughes, E., Gulcehre, C., Ortega, P., Strouse, D., Leibo, J. Z. & De Freitas, N. (2019). Social influence as intrinsic motivation for multi-agent deep reinforcement learning. *Proceedings of the 36th International Conference on Machine Learning*
- Jin, B. & Guo, W. (2022). Data driven modeling social media influence using differential equations. 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)
- Jin, B., Zou, M., Wei, Z. & Guo, W. (2025). How to find opinion leader on the online social network. *Applied Intelligence*, 55(624)
- Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A. & Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron*, 61(1)
- Lore, N. & Heydari, B. (2024). Strategic behavior of large language models and the role of game structure versus contextual framing. *Scientific Reports*, 14(18490)
- Lu, Y., Aleta, A., Du, C., Shi, L. & Moreno, Y. (2024). LLMs and generative agent-based models for complex systems research. *Physics of Life Reviews*, 51, 283–293

- Luo, B., Liu, D. & Wu, H.-N. (2017). Adaptive constrained optimal control design for data-based nonlinear discrete-time systems with critic-only structure. *IEEE Transactions on Neural Networks and Learning Systems*, 29(6), 2099–2111
- Meyer, S., Elswiler, D., Ludwig, B., Fernandez-Pichel, M. & Losada, D. E. (2022). Do we still need human assessors? Prompt-based GPT-3 user simulation in conversational AI. *Proceedings of the 4th Conference on Conversational User Interfaces*
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S. & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533
- Ndousse, K. K., Eck, D., Levine, S. & Jaques, N. (2021). Emergent social learning via multi-agent reinforcement learning. In M. Meila & T. Zhang (Eds.), *Proceedings of the 38th International Conference on Machine Learning*, vol. 139, (pp. 7991–8004). New York, NY: Proceedings of Machine Learning Research
- OpenAI (2022). GPT-3.5. Available at: <https://openai.com/>
- Panagopoulos, D., Perrusquia, A. & Guo, W. (2024). Selective exploration and information gathering in search and rescue using hierarchical learning guided by natural language input. *2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*
- Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P. & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*
- Park, J. S., Popowski, L., Cai, C., Morris, M. R., Liang, P. & Bernstein, M. S. (2022). Social simulacra: Creating populated prototypes for social computing systems. *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*
- Sekulić, I., Aliannejadi, M. & Crestani, F. (2022). Evaluating mixed-initiative conversational search systems via user simulation. *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*
- Sert, E., Bar-Yam, Y. & Morales, A. J. (2020). Segregation dynamics with reinforcement learning and agent based modeling. *Scientific Reports*, 10(1), 11771
- Shepherd, P. & Goldsmith, J. (2020). A reinforcement learning approach to strategic belief revelation with social influence. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(10), 13734–13735
- Siahkali, F., Samadi, S. & Kebriaei, H. (2024). Towards opinion shaping: A deep reinforcement learning approach in bot-user interactions. *arXiv:2409.11426*
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T. & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489
- Wang, C., Liu, Y., Gao, X. & Chen, G. (2021). A reinforcement learning model for influence maximization in social networks. In C. S. Jensen, E.-P. Lim, D.-N. Yang, W.-C. Lee, V. S. Tseng, V. Kalogeraki, J.-W. Huang & C.-Y. Shen (Eds.), *Database Systems for Advanced Applications*, (pp. 701–709). Cham: Springer International Publishing
- Watkins, C. J. & Dayan, P. (1992). Q-learning. *Machine Learning*, 8, 279–292
- Zhou, H., Zeng, D. & Zhang, C. (2009). Finding leaders from opinion networks. *2009 IEEE International Conference on Intelligence and Security Informatics*
- Zou, M., Guo, W. & Chu, K.-F. (2024). Rewiring complex networks to achieve cluster synchronization using graph convolution networks with reinforcement learning. *IEEE Transactions on Network Science and Engineering*, 11(5), 4293–4304