SPARK: Simulating the Co-evolution of Stance and Topic Dynamics in Online Discourse with LLM-based Agents

Bowen Zhang[1†], **Yi Yang**[2†], **Fuqiang Niu**[3], **Xianghua Fu**[1], **Genan Dai**[1*], **Hu Huang**[3*]
[1]College of Big Data and Internet, Shenzhen Technology University, Shenzhen, China
[2]College of Applied Science, Shenzhen University, Shenzhen, China
[3]School of Cyber Science and Technology,
University of Science and Technology of China, Hefei, China

Abstract

Topic evolution and stance dynamics are deeply intertwined in online social media, shaping the fragmentation and polarization of public discourse. Yet existing dynamic topic models and stance analysis approaches usually consider these processes in isolation, relying on abstractions that lack interpretability and agent-level behavioral fidelity. We present stance and topic evolution reasoning framework (SPARK), the first LLM-based multi-agent simulation framework for jointly modeling the co-evolution of topics and stances through natural language interactions. In SPARK, each agent is instantiated as an LLM persona with unique demographic and psychological traits, equipped with memory and reflective reasoning. Agents engage in daily conversations, adapt their stances, and organically introduce emergent subtopics, enabling interpretable, fine-grained simulation of discourse dynamics at scale. Experiments across five real-world domains show that SPARK captures key empirical patterns—such as rapid topic innovation in technology, domain-specific stance polarization, and the influence of personality on stance shifts and topic emergence. Our framework quantitatively reveals the bidirectional mechanisms by which stance shifts and topic evolution reinforce each other, a phenomenon rarely addressed in prior work. SPARK provides actionable insights and a scalable tool for understanding and mitigating polarization in online discourse. Code and simulation resources will be released after acceptance.

1 Introduction

Online social media has fundamentally transformed public discourse, enabling topics to rapidly proliferate and diversify through large-scale discussions. Rather than remaining static, many topics undergo continuous evolution in online discussions. As users repeatedly share, reinterpret, and debate content, topics may shift in meaning, give rise to related subtopics, or even merge with other themes over time. Understanding such topic evolution is central to analyzing the fragmentation of debates, the spread of misinformation, and the changing landscape of public stances in digital communities (Blei and Lafferty, 2006; Wu et al., 2024a; Tucker et al., 2018). For instance, discussions on gun control in the U.S. frequently branch into debates over constitutional rights or political division (Spitzer, 2020). Similar patterns appear across domains such as gene editing, self-driving cars, and climate change (Meyer and Vergnaud, 2023; Spoehel and Banik, 2020; Treen et al., 2020).

A variety of dynamic topic modeling approaches have been developed to capture how topics evolve over time. These methods can be broadly categorized as probabilistic dynamic topic models—which extend classical Latent Dirichlet Allocation (LDA) to temporal or networked settings using variational inference or Gibbs sampling (Blei and Lafferty, 2006; Griffiths and Steyvers, 2004; Wang et al., 2012)—and neural dynamic topic models, which leverage deep neural networks to model the semantic evolution of document collections and their network structure (Wu et al., 2024a,b).

However, probabilistic dynamic topic models typically encode topics and documents as static or smoothly changing distributions over words, making it difficult to capture fine-grained semantic shifts, user-level reasoning, or interactive stance dynamics. Neural dynamic topic models, while more expressive in modeling text and network structure, still represent topics as latent vectors and lack the interpretability and explicit reasoning needed for tracking individual attitudes and fine-grained topic evolution. Moreover, existing models generally treat topic content as abstract distributions rather than natural language, limiting their ability to simu-

* † These authors contributed equally.
Corresponding authors: {daigenan@sztu.edu.cn, huanghu@ustc.edu.cn}

Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pages 23061–23073
November 4-9, 2025

late or explain the co-evolution of topic and stance dynamics at the language level.

Recent research highlights that topic evolution and user stance are deeply intertwined. On one hand, opposing stances within a community frequently spark viewpoint clashes that determine the direction and trajectory of topic evolution, as demonstrated by analyses of contentious events such as Messi's visit to Hong Kong (Huang et al., 2024). On the other hand, the emergence of new subtopics can subsequently trigger shifts in user stances and facilitate changes in attitudes over time (Lorenz-Spreen et al., 2023). This dynamic, bidirectional relationship underscores the necessity of investigating co-evolutionary mechanisms between topic evolution and stance dynamics. Advancing our understanding of these mechanisms is essential for developing more effective strategies to guide public opinion online and mitigate the risks associated with extreme polarization.

To address these challenges, we propose stance and topic evolution reasoning framework (SPARK), the first (to our knowledge) large language model (LLM-based) multi-agent simulation framework for modeling the coordinated evolution of topic and stance dynamics in online discourse. In SPARK, each agent is instantiated as an LLM agent endowed with a unique persona, enabling agents to dynamically update their stances, autonomously generate and discuss new subtopics, and participate in complex processes of topic evolution—including branching, convergence, and transformation of discussion themes. By leveraging the reasoning and reflective capabilities of LLM agents, SPARK provides interpretable, fine-grained simulations of how topic evolution and stance change co-evolve over time. The framework supports large-scale, agent-level tracking and analysis, offering novel insights into the mechanisms that drive topic and stance dynamics in social media.

Concretely, each agent in SPARK is initialized with a unique persona—including attributes such as age, background, and personality traits—and participates in daily conversational rounds with randomly selected peers. During these interactions, agents reflect on and update their stances, and may propose or engage with new derivative topics based on their evolving viewpoints. Agents are equipped with both short-term and long-term memory modules to record daily exchanges and retain broader context, as well as a reflective reasoning process to better emulate human-like decision making and topic engagement. This setup enables the simulation to closely mirror the dynamic, intertwined evolution of topics and stances observed in real-world social networks.

Our main contributions are as follows: (1) We introduce SPARK, the first LLM-based multi-agent simulation framework for modeling the coordinated evolution of topics and stances in online social discourse. (2) SPARK enables agent-level, interpretable, and fine-grained simulation of topic and stance dynamics, supporting the analysis of co-evolutionary mechanisms with high fidelity. (3) Through comprehensive simulation experiments, we demonstrate that SPARK closely reproduces real-world evolution patterns, revealing, for example, that technology-related topics evolve more rapidly than science or healthcare topics, and that certain agent traits are associated with a higher likelihood of stance change and topic innovation. (4) Our framework provides actionable insights for online discourse management and polarization mitigation, offering a scalable tool for both research and practical intervention.

2 Related Work

Dynamic Topic Model. Dynamic topic modeling has evolved from early probabilistic frameworks to recent neural and representation learning approaches. The classical Dynamic Topic Model (Blei and Lafferty, 2006), based on Latent Dirichlet Allocation (LDA) (Blei et al., 2003), models topic evolution via state-space methods, with further extensions to continuous time (Wang et al., 2012) and nonparametric settings (Caron et al., 2012). Neural dynamic topic models, such as DETM (Dieng et al., 2019), leverage variational autoencoders to capture non-linear topic transitions, while subsequent works incorporate temporal document networks (Zhang and Lauw, 2022; Cvejoski et al., 2023), contrastive learning (Wu et al., 2024a), and pretrained Transformers (Wu et al., 2024b; Zhang et al., 2025) to enhance topic tracking and semantic structure discovery. Despite these advances, prior work models topic dynamics but largely overlooks stance and interaction signals. We explicitly capture their interplay to better explain topic evolution.

Stance Dynamics Model. Stance dynamics concerns how users' attitudes toward specific topics evolve over time on social media. Early approaches extended topic models to jointly track topics and associated stances, such as the Dynamic

Joint Sentiment-Topic model (He et al., 2014) and dynamic LDA variants for trend detection (Sasaki et al., 2014). Subsequent work explored user-level stance classification via active learning (Volkova and Van Durme, 2015) and applied neural architectures—incorporating attention mechanisms and recurrent networks—to capture temporal stance patterns from tweet histories and neighbor interactions (Chen et al., 2018; Zhu et al., 2020). Recent advances employ clustering (Azmi and Al-Ghadir, 2024) and pretrained Bert model (Unlu et al., 2025) for dynamic stance identification and tracking in large-scale datasets. However, most existing models rely on structured representations and numerical inference, limiting their ability to simulate or interpret stance evolution through natural language interactions. In contrast, our work introduces an LLM-based simulation framework, enabling agent-driven, text-based modeling of stance and topic co-evolution.

LLM-based Agents for Social Simulation. The use of LLMs as generative agents in social simulation is an emerging research direction that has demonstrated remarkable capabilities in modeling complex human behaviors (Park et al., 2023; Kaiya et al., 2023; Li et al., 2023; Zhang et al., 2024; Guo et al., 2024; Liu et al., 2024b). LLM-based agents excel at producing natural language outputs and have been shown to simulate phenomena such as trust dynamics (Xie et al., 2024), generate social media content indistinguishable from human writing (Park et al., 2022), and reproduce opinion dynamics and echo chamber effects through multi-agent interactions (Wang et al., 2025; Cau et al., 2025; Gu et al., 2025). These advances highlight the potential of LLM agents for modeling group-level social processes. To the best of our knowledge, our work is the first to leverage LLM-based agents for simulating the co-evolution of topic evolution and stance dynamics in online discourse.

3 Method

3.1 Problem Formulation

Formally, let $\mathcal{A} = \{a_1, ..., a_N\}$ denote N agents, each initialized with persona p_i and a stance s_i^0 on the original topic O. The simulation runs for T discrete steps; at each step t, every agent interacts with c peers, discussing the current set of topics $\mathcal{T}^t$. Agents update their stances s_i^t and may propose new subtopics, expanding the topic tree $\mathcal{G}^t$ to reflect ongoing topic evolution. Each agent maintains short-term and long-term memory for stance reasoning. The outputs include the evolving topic tree, agent stance trajectories, and aggregate measures of topic diversity and polarization. Our objective is to capture and analyze the bidirectional dynamics between topic evolution and stance change.

3.2 Framework Overview

We propose a multi-agent simulation framework to model the co-evolution of topic evolution and stance dynamics in online discourse. As shown in Figure 1, a population of N LLM-based agents—each with a unique persona—interacts over discrete rounds. In each round, agents engage in natural language conversations, update their stances, and may introduce new subtopics, driving the growth of a dynamic topic tree. Dual memory modules enable agents to integrate both recent and historical experiences, supporting nuanced, context-aware stance shifts. This framework provides a flexible testbed for analyzing the intertwined dynamics of topic and stance evolution at scale.

3.3 Stance-Aware Role Agent (SARA)

The Stance-Aware Role Agent (SARA) module models each individual in the simulation as an autonomous LLM-based agent that expresses, updates, and reasons about its stance through natural language interaction. SARA integrates persona initialization, a dual memory architecture, and text-driven stance reasoning, enabling nuanced and interpretable simulation of social behaviors.

Agent Persona Initialization. Each agent a_i is initialized with a distinct persona p_i, capturing demographic and psychological attributes such as name, age, education level, and personality traits. We adopt the Big Five trait model (Barrick and Mount, 1991) to sample personality dimensions, reflecting real-world individual diversity. These attributes influence the agent's initial stance s_i^0 on the original topic and modulate its susceptibility to conversational influence throughout the simulation.

Dual Memory System. To emulate human-like memory processes, SARA maintains two memory modules for each agent: short-term memory m_i^s, which records and summarizes the agent's daily interactions, and long-term memory m_i^l, which accumulates and compresses historical experiences across rounds. At the end of each simulation day,

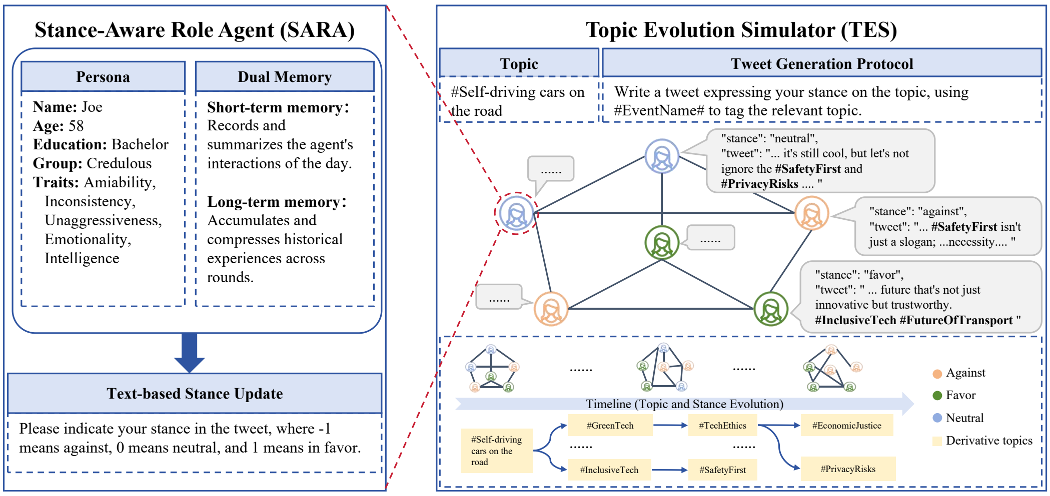

Figure 1: Our framework simulates topic evolution by equipping each agent with role-based decision-making capabilities. Stance Role-aware Agents interact within the Topic Evolution Simulator to update their stances, where their role identities shape how they engage with the topic.

short-term memory is summarized using a prompt-driven LLM call, then integrated into long-term memory via a separate update prompt. This design balances scalability with historical fidelity, allowing agents to base their stance updates on both recent and cumulative context. We use the following prompt for memory summarization:

> m_i^s prompt: The topic is *[topic]*. Here are the opinions you've heard: *[opinions]*. Briefly summarize these opinions and their stances.

> m_i^l prompt: Previous long term memory: *[long-memory]*. Today's short-term summary: *[short-memory]*. Please update the long-term memory by integrating today's summary, ensuring continuity and adding any new insights. Return only the updated long-term memory.

Text-based Stance Update. Instead of scalar or categorical stance variables, SARA represents and updates each agent's stance using natural language—specifically, tweet-format statements. At the end of each day, the agent generates a tweet that reflects its current attitude toward the topic, naturally incorporating recent interactions, personal traits, and trending subtopics. The tweet-generation prompt encourages succinct, context-aware, and human-like expression, and is annotated with a stance label (-1 for against, 0 for neutral, 1 for in favor). This approach enables interpretable, fine-grained tracking of stance evolution, and closely aligns with real-world social media behavior. The prompt for tweet generation is given below:

> In this role play, act as a real social media user. Write a tweet reflecting your opinion, mentioning related discussion and popular topics or memes. Indicate your stance, where -1 is against, 0 is neutral, and 1 is for.

3.4 Topic Evolution Simulator (TES)

TES orchestrates the dynamic development and dissemination of discussion topics across the agent population. TES models how topics diversify, branch, and propagate in response to ongoing multi-agent interactions, thereby capturing the organic formation and evolution of topic communities in online discourse.

Topic Evolution and Propagation. At each simulation round, agents are exposed to the current set of active topics and may introduce novel subtopics based on the context of their conversations. Specifically, whenever an agent's generated tweet includes a previously unseen hashtag or subtopic, TES expands the topic tree $\mathcal{G}^t$ by adding a new subtopic node corresponding to this emergent topic. This mechanism allows for organic topic diversification and mirrors the spontaneous emergence of sub-communities observed in real-world social media. The propagation of topics is further influenced by the frequency and reach of associated hashtags within the agent population. TES tracks the diffusion trajectory of each subtopic, enabling quantification of topic popularity, lifespan, and the degree of topic diversification over time.

Tweet Generation Protocol To ensure realistic and structured topic evolution, we standardize the content generation process for all agent tweets. Each agent, when updating its stance, is prompted to generate a tweet that (1) reflects its current attitude, (2) references relevant discussions or trending memes, and (3) includes only event-specific hashtags in the #EventName# format. The prompt explicitly encourages agents to integrate their prior stance and recent conversational context, while naturally incorporating or responding to new subtopics as they arise. This unified protocol enables agents to participate in coherent, evolving discussions, and provides a controlled yet flexible way to analyze how new topics are introduced and disseminated. To standardize agent tweets, our unified protocol is as follows:

> Provide the content of a tweet you might write reflecting your stance, introducing relevant discussion as appropriate, and referencing popular topics or memes.Your tweet text, including specific event hashtags in #EventName# format.

The complete prompt template can be found in Appendix A.

3.5 Simulation Algorithm

Algorithm 1 summarizes the overall simulation workflow. Each simulation day proceeds as follows: (1) agents are randomly paired for topic-centered conversations; (2) each agent aggregates daily interactions in short-term memory and updates long-term memory; (3) agents update their stance and generate a tweet; (4) TES updates the topic tree based on emergent hashtags and tracks topic propagation. This algorithm enables systematic simulation and analysis of the intertwined evolution of topics and stances, supporting downstream studies on polarization, echo chamber formation, and topic diffusion dynamics.

Algorithm 1 SPARK Simulation Workflow

Require: Agent pool $\mathcal{A}$, original topic O, simulation length T, daily contacts c

Initialize personas and initial stances for all $a_i \in \mathcal{A}$
Initialize topic tree $\mathcal{G}^0 = \{O\}$
for $t = 1$ to T **do**
  for each agent a_i **do**
    Sample c peers for conversation
    Engage in topic-centered discussions; store in m_i^s
  end for
  for each agent a_i **do**
    Summarize m_i^s (short-term) via LLM prompt
    Update m_i^l (long-term) by integrating daily summary
    Generate stance-update tweet via prompt
    Annotate tweet with stance label
  end for
  TES updates topic tree $\mathcal{G}^t$ from new hashtags in tweets
  Track topic propagation and stance trajectories
  Reset all m_i^s for next round
end for
return Topic tree evolution, stance trajectories, polarization/diversity metrics

4 Experiments

4.1 Implementation Details

We implement our SPARK simulation framework in Python, leveraging the Mesa library (Kazil et al., 2020) for agent-based modeling. All LLM calls are performed via the DeepSeek-V3-0324 API.[1] Each agent's persona is initialized with a randomly sampled name (from the names-dataset), an age (uniformly sampled in [18, 64]), and Big Five personality traits (Barrick and Mount, 1991), where each trait is independently assigned as positive or negative with equal probability. We simulate a population of 108 agents, substantially exceeding the scale of previous LLM-agent simulations (Wang et al., 2024; Liu et al., 2024b,a). Agents interact over five diverse topic domains: politics, technology, society, medicine, and science. All random processes are seeded for reproducibility. Experiments are conducted on an Intel Xeon CPU with 64GB RAM. Source code and configuration files will be released to support replicability.

4.2 Metrics

To quantitatively evaluate stance and topic evolution in our simulation, we adopt two sets of established metrics from prior work (Chuang et al., 2023; Dieng et al., 2019). For stance dynamics, following (Chuang et al., 2023), we measure Stance Bias (OB), defined as the average stance across all agents at the final time step: $OB = \frac{1}{N}\sum_{i=1}^{N} s_i^T$, where s_i^T denotes the stance of agent i at the final time step. A high absolute OB indicates strong group bias toward one direction, while values near zero reflect overall neutrality. We also report Stance Diversity (OD), calculated as the stan-

[1] https://platform.deepseek.com/

Settings	Original				Credulous				Skeptical				Consistency			
	OB	OD	TC	TD	OB	OD	TC	TD	OB	OD	TC	TD	PY	PN	IY	IN
Politics	1.00	0.00	0.23	0.14	1.00	0.00	0.25	0.49	1.00	0.00	0.25	0.31	0.67	0.79	0.67	0.79
Technology	-0.20	0.97	0.23	0.22	-0.03	1.00	0.24	0.75	-0.39	0.92	0.24	0.64	0.57	0.53	0.57	0.53
Society	0.35	0.80	0.23	0.34	0.10	0.79	0.24	0.80	0.48	0.82	0.25	0.70	0.48	0.43	0.48	0.43
Medical	0.48	0.67	0.23	0.32	0.63	0.48	0.25	0.68	0.50	0.80	0.25	0.83	0.63	0.54	0.63	0.54
Science	0.89	0.46	0.23	0.43	1.00	0.00	0.25	1.32	0.85	0.53	0.26	0.83	0.54	0.58	0.54	0.58
Avg	0.50	0.58	0.23	0.29	0.54	0.45	0.25	0.81	0.49	0.61	0.25	0.66	0.58	0.57	0.58	0.57

Table 1: Comparative analysis of stance bias (OB), stance diversity (OD), topic consistency (TC), and topic diversity (TD) for Original, Credulous, and Skeptical groups by category, with consistency to prior-day stance (PY: with novel subtopics, PN: without novel subtopics), and consistency to day-one stance with (IY) or without (IN) novel subtopics included. All consistency ratios are rounded to two decimal places.

dard deviation of agent stances at the final step: $OD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(s_i^T - OB)^2}$. Higher OD values indicate greater heterogeneity or polarization among agent stances, while lower OD suggests consensus.

For topic evolution, we consider Topic Coherence (TC), computed as the entropy of the topic distribution within each time slice: $TC^t = -\sum_k P_k^t \log P_k^t$, where P_k^t is the normalized frequency of topic k at time t. Lower TC values indicate more focused and coherent discussions, whereas higher TC reflects broader topic coverage. Additionally, we assess Topic Diversity (TD) using the Kullback-Leibler (KL) divergence between topic distributions in consecutive time slices: $TD^t = D_{KL}(P^t \| P^{t-1})$. Higher TD signifies greater novelty and change in the topics discussed over time. We report the average TC and TD across all time slices to capture both the consistency and the dynamism of topic evolution in the simulated environment.

4.3 Macro-level Analysis

This section addresses three core research questions regarding collective opinion and topic evolution in simulated agent societies:

(1) How do agent personality traits and discussion domains influence stance dynamics?

(2) How do traits and topic types affect the process of topic evolution?

(3) How do topic evolution and stance change mutually influence each other?

We systematically investigate these questions using established quantitative metrics and simulation results, as detailed below.

Personality Trait Effects on Stance. We further examine the impact of individual differences by grouping agents as “Credulous” or “Skeptical” based on their Agreeableness and Neuroticism (Oyibo and Vassileva, 2019; Widiger and Oltmanns, 2017). As shown in Table 1, Credulous agents exhibit a greater mean stance change (0.54) than Skeptical agents (0.49), while stance variance remains comparable (0.45 vs. 0.61), confirming the stability of our simulation. These results align with psychological theory, and demonstrate that personality-driven susceptibility to social influence can be robustly reproduced and quantified in large-scale LLM-agent settings.

Domain Effects on Stance Dynamics. Stance dynamics display marked variation across domains. In Politics and Science, agents converge to a single extreme, as indicated by high stance bias (OB = 1.00 and 0.89) and low stance diversity (OD = 0.00 and 0.46), reflecting strong consensus or polarization. In contrast, Technology (OB = -0.20) and Society (OB = 0.35) domains exhibit more moderate bias and maintain greater diversity (OD = 0.97 and 0.80), suggesting a broader spectrum of opinions. Medical presents intermediate values (OB = 0.48, OD = 0.67). On average, stance bias and diversity are 0.50 and 0.58, indicating a tendency toward moderate group bias with coexistence of consensus and heterogeneity. These results highlight the significant influence of topic type on collective stance formation, and systematically confirm established findings in opinion dynamics at previously unexplored simulation scale. The detailed information on the number of derived topics in all domains and the stance change curves can be found in Appendix B and Table 1.

Trait and Domain Effects on Topic Evolution. Topic consistency (TC) is uniformly high across domains (TC = 0.23), indicating focused and coherent discussions within each time slice. However, topic novelty (TD) varies: Science (0.43), Society (0.34), and Medical (0.32) display more dynamic topic shifts, while Politics remains more stable (0.14).

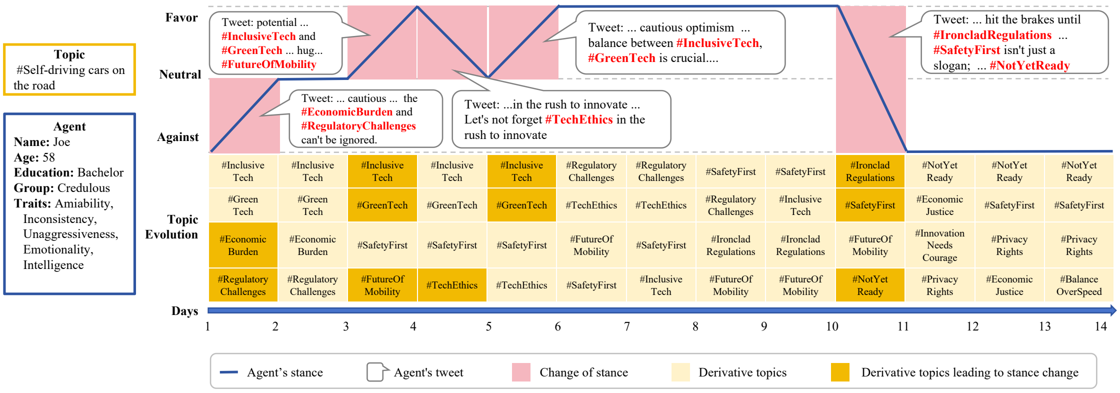

Figure 2: A micro-level case study of topic evolution and stance change. Agent Joe changes his stance in response to the emergence of new subtopics. The emergent subtopics associated with Joe's stance changes are highlighted in dark color.

Across all domains, Credulous agents consistently drive higher topic novelty than Skeptical agents (average TD: 0.81 vs. 0.66), suggesting that personality not only affects stance but also fosters greater topic innovation and dynamism. Notably, this provides large-scale simulation evidence for the role of personality in social information innovation.

Interplay Between Topic Evolution and Stance Change. A time-series analysis reveals a strong positive correlation (Pearson $r = 0.88$) between stance change and the emergence of new topics, indicating that shifts in stance actively promote topic generation. Moreover, group stance consistency drops significantly on days when new topics appear (e.g., in Politics, from 0.79 to 0.67), revealing mutual reinforcement between topic innovation and stance diversity. This coupling effect demonstrates our framework's ability to capture complex, bidirectional dynamics between collective opinions and evolving discourse structures—a relationship rarely quantitatively demonstrated in previous LLM-based simulations.

Overall, our findings systematically reproduce and quantify theoretical patterns of domain and personality effects on stance and topic dynamics, and further reveal their intricate mutual reinforcement in simulated social environments.

4.4 Micro-level Analysis

Case Study. To illustrate the dynamic interplay between topic exposure and stance evolution at the individual level, we present the trajectory of an agent ("Joe") engaging in discussions on self-driving cars (see Figure 2).

Initially, Joe holds an opposing stance, voicing concerns about #Self-driving cars on the road. As derivative topics such as #EconomicBurden and #RegulatoryChallenges emerge, Joe's attitude shifts to neutral, reflecting a more balanced consideration of both risks and benefits. With the continued emergence of topics such as #GreenTech, #TechEthics, and #FutureOfMobility, Joe enters a phase of hesitation, marked by fluctuating views as new arguments are introduced (Days 3 to 6). Sustained positive discussions—especially around #InclusiveTech and #GreenTech—eventually move Joe toward a supportive stance. However, the emergence of critical issues such as #IroncladRegulations and #NotYetReady triggers a marked reversal, and Joe adopts an opposed position as further concerns (e.g., #PrivacyRights, #EconomicJustice) come to the forefront. This case highlights how our framework captures nuanced stance dynamics, revealing how evolving topics and agent traits jointly drive complex opinion trajectories over time.

Mechanism Illustration. This individual-level trajectory highlights a dynamic, bidirectional mechanism underpinning stance-topic co-evolution. The emergence of new hashtags introduces fresh perspectives and concerns, which not only diversify the topical landscape but also act as catalysts for shifts in Joe's stance. Conversely, as Joe's viewpoint evolves—becoming more critical or supportive—he actively engages with, and helps propagate, newly emergent topics. This feedback loop creates a self-reinforcing cycle: topic innovation stimulates stance change, while evolving stances fuel the adoption and spread of additional topics.

Implications and Connection to Macro-level

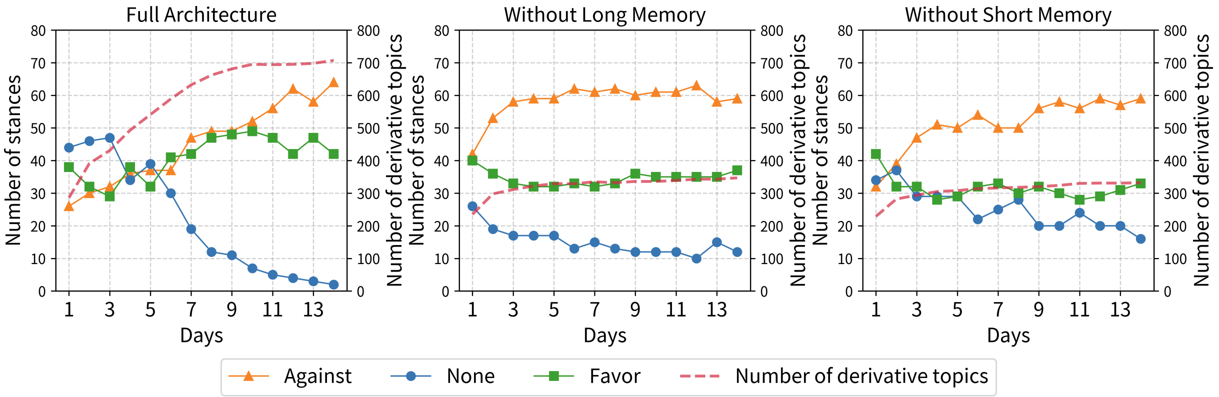

Figure 3: Comparison of stance distribution and the number of emergent subtopics for the technology domain. In the Full Architecture setting, as topics evolve and the number of emergent subtopics increases, the number of neutral agents decreases, while the number of agents with clear stances (in favor or against) increases.

Trends. The micro-level dynamics observed in Joe's discussion trajectory mirror the aggregate patterns identified in our macro-level analysis: greater stance diversity is associated with increased topic novelty and variability. This case study demonstrates how individual agents, through their responsiveness to new topics and willingness to adapt their stances, serve as primary drivers of collective discourse innovation. The findings thus reinforce the validity of our simulation framework and provide a concrete mechanistic explanation for the observed mutual reinforcement between topic evolution and stance change at the population level.

5 Analysis and Discussion

Persistent Stance Holders in Topic Dynamics. Our analysis reveals that a subpopulation of agents consistently maintain their initial stances throughout the simulation, indicating that derivative topics exert only limited influence on these individuals. Further examination shows that these persistent stance holders span a range of ages and education levels, but share common personality characteristics—most notably, low neuroticism. This finding is consistent with classic psychological literature (John et al., 2010), which links low neuroticism to greater resistance to external influence and opinion change. These results suggest that policy interventions or public discourse strategies aimed at shifting collective stances may benefit from tailoring approaches to specific personality profiles, rather than adopting uniform strategies. These results have important implications for real-world applications such as social media governance and public opinion management, as they suggest that effective stance-shifting interventions require personalized, trait-aware strategies rather than uniform or population-level approaches.

Ablation Study. To assess the contribution of key model components, we conduct an ablation study on a representative science topic, systematically removing long-term memory and short-term memory modules, as shown in Figure 3. When long-term memory is ablated, agents rely solely on short-term interactions, which proves insufficient to sustain meaningful topic evolution and stance change; interactions stagnate and topic diversity diminishes. In contrast, removing short-term memory impairs the integration of recent information, resulting in long-term memory that passively aggregates daily stances without enabling genuine reflection or consolidation. This leads to a monotonous evolution of both derivative topics and stances. Quantitatively, when short-term memory is ablated, stance diversity—measured by the metric from (Chuang et al., 2023)—declines to approximately three-quarters of that in the full model (see Appendix C for details). These findings underscore that memory components are essential for maintaining realistic and diverse stance dynamics.

6 Conclusion

We introduce SPARK, the first LLM-based multi-agent simulation framework that jointly models the co-evolution of topics and stances in online discourse. By instantiating each agent as an LLM persona with memory and reflective reasoning, SPARK enables interpretable, fine-grained simulations of large-scale discourse dynamics. Our experiments across five domains reveal not only key empirical regularities—including rapid topic innovation, domain-specific polarization, and personality-driven stance shifts—but also, crucially, the bidirectional mechanisms by which topic evolution and stance change reinforce each other, a phenomenon rarely addressed in prior work. SPARK offers a scalable platform for agent-based social modeling, providing both theoretical insight and practical tools to study and mitigate online polarization. In future work, we aim to incorporate richer network structures and external information flows, further advancing the study of complex social discourse.

Limitations

This study only investigates the impact of Big Five personality traits and topic types on the co-evolution of topic-stance synergy, without incorporating other potential factors that may influence the evolution of social networks. In addition, the current simulation environment is relatively simplified and does not fully capture the real-world complexity of mainstream social platforms such as Facebook and Twitter. Therefore, the applicability of our findings to larger-scale, more diverse, and more realistic social network environments still requires further validation. In future work, we plan to conduct simulations in environments that more closely resemble real social platforms, in order to enhance the practical significance and generalizability of our results.

Ethics Statement

This study utilizes a large language model (LLM)-based multi-agent simulation approach to investigate the co-evolution of topics and stances in online social media. No real human participants or personal data were involved during the research process. All agents are synthetic personas, and their demographic and psychological traits are designed solely for modeling social phenomena in a controlled and ethical environment. The design of personality traits aims to better understand discourse dynamics such as polarization and topic innovation, and does not imply endorsement of any specific viewpoint or behavior. The code and simulation resources are intended solely for research purposes, and we ensure that all usage complies with their intended purpose.

Acknowledgements

This research is supported by the National Natural Science Foundation of China No. 62306184, Natural Science Foundation of Top Talent of SZTU (grant no. GDRC202320), and Shenzhen Science and Technology Program (No. JCYJ20240813113218025).

References

Aqil M Azmi and Abdulrahman I Al-Ghadir. 2024. Using twitter as a digital insight into public stance on societal behavioral dynamics. *Journal of King Saud University-Computer and Information Sciences*, 36(5):102078.

Murray R Barrick and Michael K Mount. 1991. The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, 44(1):1–26.

David M Blei and John D Lafferty. 2006. Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*, pages 113–120.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

François Caron, Manuel Davy, and Arnaud Doucet. 2012. Generalized polya urn for time-varying dirichlet process mixtures. *arXiv preprint arXiv:1206.5254*.

Erica Cau, Valentina Pansanella, Dino Pedreschi, and Giulio Rossetti. 2025. Language-driven opinion dynamics in agent-based simulations with llms. *arXiv preprint arXiv:2502.19098*.

Chengyao Chen, Zhitao Wang, and Wenjie Li. 2018. Tracking dynamics of opinion behaviors with a content-based sequential opinion influence model. *IEEE Transactions on Affective Computing*, 11(4):627–639.

Yun-Shiuan Chuang, Agam Goyal, Nikunj Harlalka, Siddharth Suresh, Robert Hawkins, Sijia Yang, Dhavan Shah, Junjie Hu, and Timothy T Rogers. 2023. Simulating opinion dynamics with networks of llm-based agents. *arXiv preprint arXiv:2311.09618*.

Kostadin Cvejoski, Ramsés J Sánchez, and César Ojeda. 2023. Neural dynamic focused topic model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 12719–12727.

Adji B Dieng, Francisco JR Ruiz, and David M Blei. 2019. The dynamic embedded topic model. *arXiv preprint arXiv:1907.05545*.

Thomas L Griffiths and Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl_1):5228–5235.

Chenhao Gu, Ling Luo, Zainab Razia Zaidi, and Shanika Karunasekera. 2025. Large language model driven agents for simulating echo chamber formation. *arXiv preprint arXiv:2502.18138*.

Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V Chawla, Olaf Wiest, and Xiangliang Zhang. 2024. Large language model based multi-agents: A survey of progress and challenges. *arXiv preprint arXiv:2402.01680*.

Yulan He, Chenghua Lin, Wei Gao, and Kam-Fai Wong. 2014. Dynamic joint sentiment-topic model. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(1):1–21.

Shan Huang, Tao Wang, Yuanhan Xie, Jiayuan Sun, Lifang Li, Weishan Zhang, and Fei-Yue Wang. 2024. Opposing stance in topic evolution: A case analysis of messi's visit to hong kong. *IEEE Transactions on Computational Social Systems*.

Oliver P John, Richard W Robins, and Lawrence A Pervin. 2010. *Handbook of personality: Theory and research*. Guilford Press.

Zhao Kaiya, Michelangelo Naim, Jovana Kondic, Manuel Cortes, Jiaxin Ge, Shuying Luo, Guangyu Robert Yang, and Andrew Ahn. 2023. Lyfe agents: Generative agents for low-cost real-time social interactions. *arXiv preprint arXiv:2310.02172*.

Jackie Kazil, David Masad, and Andrew Crooks. 2020. Utilizing python for agent-based modeling: The mesa framework. In *Social, Cultural, and Behavioral Modeling: 13th International Conference, SBP-BRiMS 2020, Washington, DC, USA, October 18–21, 2020, Proceedings 13*, pages 308–317. Springer.

Chao Li, Xing Su, Haoying Han, Cong Xue, Chunmo Zheng, and Chao Fan. 2023. Quantifying the impact of large language models on collective opinion dynamics. *arXiv preprint arXiv:2308.03313*.

Yuhan Liu, Xiuying Chen, Xiaoqing Zhang, Xing Gao, Ji Zhang, and Rui Yan. 2024a. From skepticism to acceptance: Simulating the attitude dynamics toward fake news. *arXiv preprint arXiv:2403.09498*.

Yuhan Liu, Zirui Song, Xiaoqing Zhang, Xiuying Chen, and Rui Yan. 2024b. From a tiny slip to a giant leap: An llm-based simulation for fake news evolution. *arXiv preprint arXiv:2410.19064*.

Philipp Lorenz-Spreen, Lisa Oswald, Stephan Lewandowsky, and Ralph Hertwig. 2023. A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nature human behaviour*, 7(1):74–101.

Morgan Meyer and Frédéric Vergnaud. 2023. The gene-edited babies controversy: Reactions in the scientific community, social media, and the press. In *Genome Editing in Biomedical Sciences*, pages 191–204. Springer.

Kiemute Oyibo and Julita Vassileva. 2019. The relationship between personality traits and susceptibility to social influence. *Computers in Human Behavior*, 98:174–188.

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.

Joon Sung Park, Lindsay Popowski, Carrie Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2022. Social simulacra: Creating populated prototypes for social computing systems. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pages 1–18.

Kentaro Sasaki, Tomohiro Yoshikawa, and Takeshi Furuhashi. 2014. Online topic model for twitter considering dynamics of user interests and topic trends. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1977–1985.

Robert J Spitzer. 2020. *The politics of gun control*. Routledge.

Elizabeth Spoehel and Shankar Banik. 2020. Self driving cars: All you need to know. In *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 655–658. IEEE.

Kathie M d'I Treen, Hywel TP Williams, and Saffron J O'Neill. 2020. Online misinformation about climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 11(5):e665.

Joshua A Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. 2018. Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)*.

Ali Unlu, Sophie Truong, Nitin Sawhney, Jonas Sivelä, and Tuukka Tammi. 2025. Tracing the dynamics of misinformation and vaccine stance in finland amid covid-19. *Information, Communication & Society*, 28(2):193–217.

Svitlana Volkova and Benjamin Van Durme. 2015. Online bayesian models for personal analytics in social media. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29.

Chenxi Wang, Zongfang Liu, Dequan Yang, and Xiuying Chen. 2024. Decoding echo chambers: Llm-powered simulations revealing polarization in social networks. *arXiv preprint arXiv:2409.19338*.

Chenxi Wang, Zongfang Liu, Dequan Yang, and Xiuying Chen. 2025. Decoding echo chambers: LLM-powered simulations revealing polarization in social networks. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 3913–3923, Abu Dhabi, UAE. Association for Computational Linguistics.

Chong Wang, David Blei, and David Heckerman. 2012. Continuous time dynamic topic models. *arXiv preprint arXiv:1206.3298*.

Thomas A Widiger and Joshua R Oltmanns. 2017. Neuroticism is a fundamental domain of personality with enormous public health implications. *World psychiatry*, 16(2):144.

Xiaobao Wu, Xinshuai Dong, Liangming Pan, Thong Nguyen, and Anh Tuan Luu. 2024a. Modeling dynamic topics in chain-free fashion by evolution-tracking contrastive learning and unassociated word exclusion. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 3088–3105, Bangkok, Thailand. Association for Computational Linguistics.

Xiaobao Wu, Thong Nguyen, Delvin Zhang, William Yang Wang, and Anh Tuan Luu. 2024b. Fastopic: Pretrained transformer is a fast, adaptive, stable, and transferable topic model. *Advances in Neural Information Processing Systems*, 37:84447–84481.

Chengxing Xie, Canyu Chen, Feiran Jia, Ziyu Ye, Shiyang Lai, Kai Shu, Jindong Gu, Adel Bibi, Ziniu Hu, David Jurgens, and 1 others. 2024. Can large language model agents simulate human trust behavior? In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Delvin Ce Zhang and Hady Lauw. 2022. Dynamic topic models for temporal document networks. In *International Conference on Machine Learning*, pages 26281–26292. PMLR.

Delvin Ce Zhang, Menglin Yang, Xiaobao Wu, Jiasheng Zhang, and Hady W Lauw. 2025. Hierarchical graph topic modeling with topic tree-based transformer. *arXiv preprint arXiv:2502.11345*.

Xiaoqing Zhang, Xiuying Chen, Yuhan Liu, Jianzhou Wang, Zhenxing Hu, and Rui Yan. 2024. Llm-driven agents for influencer selection in digital advertising campaigns. *arXiv e-prints*, pages arXiv–2403.

Lixing Zhu, Yulan He, and Deyu Zhou. 2020. Neural opinion dynamics model for the prediction of user-level stance dynamics. *Information Processing & Management*, 57(2):102031.

A Prompt Set

Here we provide an in-depth description of the prompts used in our simulation to model the dynamics of topic evolution and stance changes.

Tweet Generation Protocol and Text-based Stance Update Prompt:

> Based on the following inputs, update your stance on the *[topic]*
> 1. Previous personal Opinion: *[opinion]*
> 2. Long Memory Summary of Others' Opinions: *[long memory]*
> 3. Name: *[agent name]*
> 4. Trait: *[agent persona]*
> 5. Education level: *[agent qualification]*
>
> Keep in mind that in this role play, you are playing a real social media user. Since humans often exhibit factual bias, you should exhibit a similar tendency. This means that you are more likely to believe information that is consistent with what you already believe, and you may oppose or be convinced by information that is contrary to what you already believe. Your responses will be displayed in JSON format. Organize them as follows:
> tweet: Provide the content of a tweet you might write reflecting your stance, introducing relevant discussion as appropriate, and referencing popular topics or memes.Your tweet text, including only specific event hashtags in #EventName# format.
> stance: This indicates your stance on the message, where -1 is against, 0 is neutral, and 1 is for.
> For example: { "tweet": "my stance hasn't wavered. #SafetyFirst and #PrivacyRights are non-negotiable. The tech's potential is huge, but without IroncladRegulations, we're playing with fire.", "stance": -1, "reasoning": "The risks, including #PrivacyRisks and #JobLosses, outweigh the benefits at this stage." }

Short-Term Memory Prompt (f_m^s):

> The dicussed topic is *[topic]*. Here are the opinions you have heard so far: *[opinions]*. Summarize the opinions you have heard in a few sentences, including their stance on the topic.

Long-Term Memory Prompt (f_m^l):

	Stance Diversity
Full Architecture	**0.504**
-w/o Short Memory	0.300
-w/o Long Memory	0.309

Table 2: Memory component ablation experiment results: the higher the value, the greater the stance diversity.

> Recap of Previous Long-Term Memory: *[long memory]*. Today's Short-Term Summary: *[short memory]*. Please update long-term memory by integrating today's summary with the existing long-term memory, ensuring to maintain continuity and add any new insights or important information from today's interactions.

B Diverse Topics and Simulation Results

Figure 4 reveals the stance distribution and topic evolution in the five domains of politics, technology, society, healthcare, and science. The results show that in politics, discussions quickly become unified, creating an echo chamber effect; in technology, topics remain diverse and contentious; in society, opinions shift gradually and evenly; in healthcare, consensus forms but some diversity remains; in science, there is strong consensus but a wide range of topics. All domains show an S-shaped growth in topic consistency, with a value of 0.23, indicating common patterns in online discussions, but group opinion evolves differently across fields.

Below, we detail the topics used in our experiments on topic evolution across different subjects.

> **Technology Topic:** "#Self-driving cars on the road#"
> **Political Topic:** "#Gun control#"
> **Social Topic:** "#Naked resignation#"
> **Medical Topic:** "#Gene editing#"
> **Science Topic:** "#Global warming is a hoax#"

C Ablation Study

We also compare stance diversity between the full architecture and versions without short-term or long-term memory. As shown in Table 2, the full architecture achieves the highest stance diversity (0.504), while removing short-term or long-term memory leads to lower diversity (0.300 and 0.309). This shows that memory components help increase

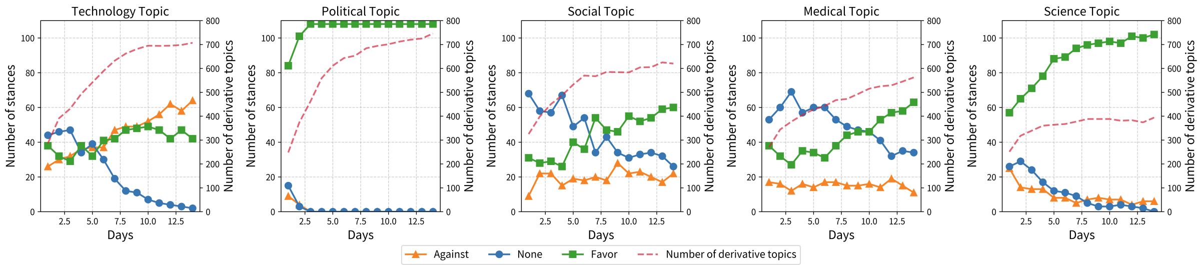

Figure 4: Tance changes and derived topic quantities across different topics.

stance diversity.