Beliefs in Motion: Simulating Opinion Dynamics via LLM-Powered Community Reactions

Xinyi Liu[1][0009−0003−4901−4796]⋆, Dachun Sun[1][0000−0003−4000−2783], Dilek Hakkani-Tür[1][0000−0001−5246−2117], and Tarek Abdelzaher[1][0000−0003−3883−7220]

University of Illinois Urbana-Champaign, 201 N Goodwin Ave, Urbana, IL 61801, USA

Abstract. Understanding how opinions shift in response to external events is essential for modeling real-world social dynamics. While traditional agent-based models offer structural insights, they lack language understanding and struggle to adapt to open, evolving information environments. Recent advances in Large Language Models (LLMs) offer a new path forward, yet assigning an LLM agent to every user is both computationally infeasible and socially noisy. In this paper, we introduce **OpinioNet**, a scalable framework that models LLM-powered community agents, each representing an ideological community. This design enables efficient, context-aware simulation while maintaining social realism. To enrich agent expressiveness and mitigate over-smoothing, OpinioNet integrates multi-level personas, including abstract ideological tags, distilled group narratives, and semantically retrieved historical posts. Additionally, simulated influencer endorsements embed representative user voices into each community's response. User-level opinions are then updated by combining external event influence, social network structure, and individual opinion inertia. Experiments show OpinioNet outperforms classical opinion dynamics models, achieving **+18.29%** Micro-F1, **+19.21%** Macro-F1 and **+39.78%** Pearson correlation, demonstrating a practical and interpretable solution for simulating ideological change at scale.

Keywords: Opinion Dynamics · Agent-Based Models · Multi-level Persona Modeling · Language-aware Agent Simulation, · Social Network Influence.

1 Introduction

How do individuals shift their opinions when confronted with major sociopolitical events? In online spaces, such dynamics unfold linguistically—users interpret, react, and adapt not in abstract numbers but through natural language. Capturing these shifts is central to understanding public discourse and modeling real-world social dynamics. Figure 1 illustrates a typical example: a user historically expresses support for Western alliances. Upon encountering a geopolitical development—paired with pressure from their community—their stance evolves into a more tempered position. This kind of language-driven ideological shift is both common and consequential, but traditional opinion dynamics models fail to capture it.

Conventional agent-based models (ABMs) such as Friedkin-Johnsen, Hegselmann-Krause, and Deffuant-Weisbuch[11, 19, 8] treat users as opinion-bearing agents within a

⋆ Reproducible code is available at: https://github.com/tracy3057/OpinionNet

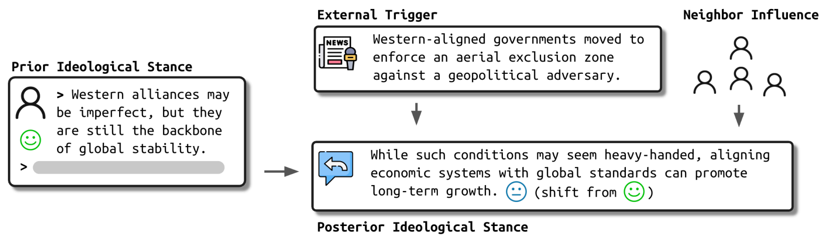


Fig. 1: Example of Language-Driven Opinion Shift.

static network. These models are scalable and interpretable but suffer two key limitations: (1) They treat opinion as a numeric value, lacking semantic grounding. (2) They ignore real-time external triggers, leading to over-simplified, over-smoothed simulations that cannot reflect complex ideological shifts. Recent breakthroughs in Large Language Models (LLMs) [4] offer a path forward: LLMs can simulate nuanced, context-aware reactions to unfolding events. However, assigning an LLM to each user is infeasible—both computationally and behaviorally—given the scale and noise of social networks.

To balance semantic richness with computational efficiency, we introduce **OpinioNet**, a novel framework that simulates opinion dynamics using LLM-powered *ideological community agents* rather than individual users. This design is both sociologically grounded and scalable: users within the same ideological community tend to exhibit aligned reactions to external events, making the community a natural unit for modeling collective opinion shifts. These agents generate language-aware responses to events and propagate their influence to users through social and ideological ties.

As shown in Figure 2, OpinioNet consists of three modules: (1) User IdeoProfiling infers users' ideological stability and assigns them to opinion communities. (2) Ideology Community Agent Modeling builds community-level agents using multi-level personas and influencer endorsements. (3) Dynamic Ideology Simulation updates user opinions by integrating community reactions, social network structure, and personal inertia.

Our key contributions are as follows:

- LLM-Powered Community Simulation: We propose OpinioNet, the first framework to model ideological opinion shifts via LLM-driven community agents, enabling semantically rich yet computationally feasible opinion dynamics simulation at scale.
- Multi-Level Persona Design: We introduce a novel multi-level persona mechanism (high-level stance, distilled narrative, history-based evidence, and influencer voice) to represent ideological communities with diverse internal semantics, mitigating over-smoothing in community-based modeling.
- Language-Aware Opinion Evolution: OpinioNet bridges classical ABMs and modern LLMs by combining symbolic social simulation (e.g., network structure, opinion inertia) with language-grounded response generation, capturing both structural and semantic opinion shifts.
- Real-World Benchmark and Evaluation Protocol: We release OpinioNet-Bench, a benchmark suite built on real-world Twitter data with ideological annotations and temporal events, along with dual evaluation metrics on (1) user-level opinion prediction and (2) community-level trend alignment.

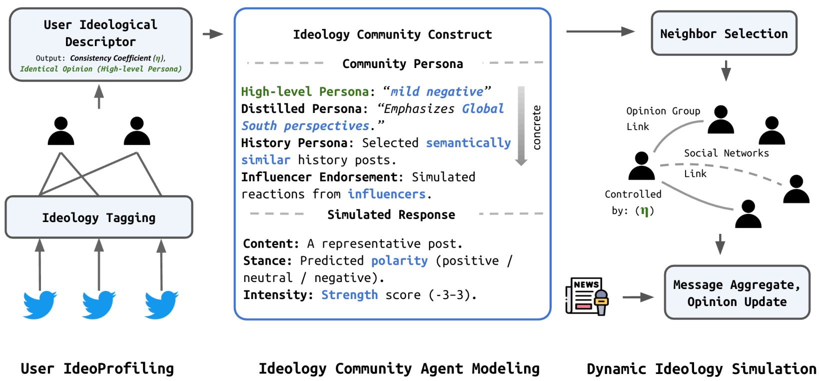


Fig. 2: Overview of OpinioNet. The framework includes: (1) User IdeoProfiling to extract ideological features; (2) Ideology Community Agent Modeling with multi-level personas and influencer responses; (3) Dynamic Ideology Simulation combining community stance, social ties, and opinion inertia.

– Strong Empirical Results: Experiments show that OpinioNet outperforms traditional opinion dynamics models across multiple tasks, improving Micro-F1 by 18.29%, Macro-F1 by 19.21% and Pearson correlation by 39.78%, demonstrating the viability of scalable, language-informed opinion simulations.

2 Related Work

2.1 Agent-Based Models for Opinion Dynamics

Agent-based models (ABMs) have long served as the backbone for simulating opinion dynamics, where individual agents update their beliefs based on interactions within a social network. Classical models—such as Friedkin-Johnsen (FJ) [11], Hegselmann-Krause (HK) [19], and Deffuant-Weisbuch (DW) [8], offering foundational frameworks for modeling opinion averaging, bounded confidence, and resistance to influence. These models capture key social phenomena like consensus formation and polarization but rely on static, rule-based mechanisms and overlook the influence of external events or language-based persuasion.

To address these gaps, extended models have incorporated richer mechanisms. The Complex Contagion model [5] emphasizes the need for multiple exposures to trigger opinion change, aligning better with empirical observations. Others, such as Martins' model [15], integrate media influence to simulate the impact of top-down information flow. However, even these enhanced models lack the ability to process or reason over language—crucial in today's discourse-rich social platforms.

In contrast, OpinioNet augments the ABM paradigm by introducing LLM-powered community agents capable of understanding, generating, and adapting to evolving linguistic inputs. This allows for simulations that are both structurally grounded and context-aware, bridging a longstanding gap between symbolic rules and semantic reasoning.

2.2 LLMs for Social Simulation

Large Language Models (LLMs) have opened new frontiers for simulating social behavior by enabling agents to interpret and generate human-like text [1, 4]. Prior work has demonstrated LLMs' capacity to model ideological framing [2], simulate online discussions [3], and coordinate multi-agent dialogue [17]. These efforts mark a departure from rule-based simulations, offering adaptability and richer context modeling.

Nonetheless, existing LLM-based simulations often focus on small-scale scenarios or short-term behaviors. They typically treat each agent as an independent LLM, which is computationally expensive and lacks scalability [13]. Moreover, most studies do not simulate the longitudinal evolution of user opinions or incorporate realistic social constraints like network structure or belief inertia [16, 22, 14].

OpinioNet addresses these limitations through a hybrid architecture: instead of assigning an LLM per user, it defines ideology-driven community agents that simulate community-level opinion trajectories. This design maintains computational feasibility while capturing long-term dynamics. Furthermore, OpinioNet integrates three key factors—opinion inertia, predisposition, and social influence—offering a theoretically grounded yet scalable method for simulating ideological shifts in response to evolving events.

3 Problem Definition

We consider a set of N users, denoted as $\mathcal{U} = \{u_1, u_2, ..., u_N\}$, who engage in online discussions on a topic by posting from a set of M posts, $\mathcal{A} = \{a_1, a_2, ..., a_M\}$. The historical posts of a user u_i are represented as $\mathcal{A}_{u_i}$. Each post $a_j \in \mathcal{A}$ carries an associated opinion score s_{a_j}, where $-3 \leq s_{a_j} \leq 3$. The sign of s_{a_j} indicates the opinion's direction, while its magnitude reflects the intensity of the stance.

Additionally, we define a set of ongoing events, represented as $E = \{e_1, e_2, ...e_k\}$, where each event e_i is expressed in natural language and may influence user opinions over time. Let $s_{u_i}^{(t)}$ denote the stance of user u_i at time window t, where $-3 \leq s_{u_i}^{(t)} \leq 3$, with similar semantics of polarity and extremeness as post-level opinions. The objective of OpinioNet is to predict the future opinion dynamics of each user $u_i \in \mathcal{U}$ over a sequence of time windows. Specifically, we aim to model the opinion trajectory:

$$S_{u_i} = \{s_{u_i}^1, s_{u_i}^2, ..., s_{u_i}^{(t)} ..., s_{u_i}^{(T)}\}$$

where t represents the time window index and T is the maximum time window considered. By modeling stance evolution in response to ongoing events and discussions, OpinioNet provides a realistic and adaptive simulation of ideological shifts in online discourse.

4 Framework of OpinioNet

The OpinioNet framework consists of three key modules: (1) User IdeoProfiling: This module labels historical posts and extracts users' opinions by analyzing their past interactions. It captures both opinion direction (positive, neutral, negative) and intensity (ranging from neutral to extreme). (2) Ideology Group Agent Modeling: This module organizes users into opinion communities with varying levels of persona concreteness, including high-level, distilled, and historical personas. It then simulates how each opinion group would respond to new events based on their historical stance and contextual influences. (3) Dynamic Ideology Simulation: This module models opinion evolution by incorporating three factors: personal opinion inertia (resistance to change), opinion predisposition (initial biases), and social network influence (interactions with others). It updates user opinions dynamically as they engage with new information and social interactions.

4.1 User IdeoProfiling

Understanding a user's ideological stance requires aggregating both their expressed opinions and their temporal consistency. We formalize this through two sequential stages: ideology tagging at the post level and ideological descriptor estimation at the user level.

Ideology Tagging. Given a user-generated post $a_j \in \mathcal{A}$, we infer its latent opinion label using a prompt-based LLM classifier $\mathbf{LLM}(\cdot, P_l)$, which maps raw text into one of seven fine-grained opinion categories: *mild/moderate/extreme positive, neutral, mild/moderate/extreme negative*. This design follows psychological theories of opinion valence and extremity, where attitude strength is considered orthogonal to polarity [18]. We convert each label into a scalar stance score $s_{a_j} \in [-3, +3]$, capturing both direction and intensity:

$$s_{a_j} = \mathbf{LLM}(a_j, P_l) \tag{1}$$

User Ideological Descriptor. For each user $u_i \in \mathcal{U}$, we aggregate their historical post opinions $\mathcal{A}_{u_i}$:

– **Consistency Coefficient** η_i: quantifies ideological stability as the relative dominance of consistent opinion direction across a user's history. Inspired by notions of opinion inertia in social psychology [10], we compute:

$$\eta_i = \frac{\sum_{a_j \in \mathcal{A}_{u_i}} \left| \mathbb{1}(s_{a_j} > 0) - \mathbb{1}(s_{a_j} < 0) \right|}{|\mathcal{A}_{u_i}|} \tag{2}$$

– **Community Assignment** C_{u_i}: reflects the user's average stance and determines their ideological group membership. The value is rounded to the nearest integer in $[-3, 3]$ to match the high-level persona classes:

$$C_{u_i} = \text{round}\left(\frac{1}{|\mathcal{A}_{u_i}|} \sum a_j \in \mathcal{A}_{u_i} s_{a_j} \right) \tag{3}$$

This profiling step provides a principled and scalable way to map users to ideology communities, serving as the foundation for community-level opinion simulation.

4.2 Ideology Community Agent Modeling

To model how ideological communities respond to external events, we build multi-level persona representations that reflect both stable group traits and context-specific narratives (Figure 3). These include a High-level Persona (shared stance), Distilled Persona (dominant rhetorical tone), History Persona (relevant past experiences), and Influencer Endorsement (elite signals). This structure captures both group coherence and contextual adaptability, drawing on social identity theory [21] and discourse-centered models of belief formation [9].

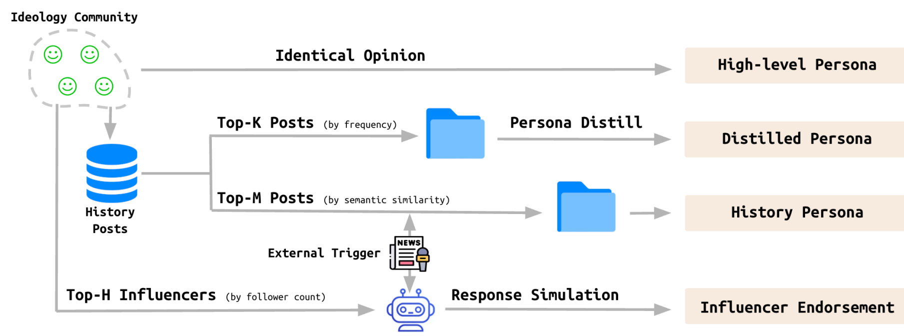


Fig. 3: Generation of Community Personas and Endorsements in Response to External Triggers.

High-level Persona. We begin by clustering users into seven discrete ideological groups according to their average post-level opinion labels (e.g., *mild positive*, *extreme negative*), reflecting dimensions of polarity and extremity. This coarse-grained categorization provides a foundational symbolic representation H_c of community c's ideological orientation.

Distilled Persona. While the high-level persona captures a community's categorical stance, it lacks narrative depth. To address this, we synthesize a distilled persona D_c for each community by prompting the LLM to summarize its top-K most representative posts (selected by frequency or engagement). The resulting summary reflects core beliefs, tone, and rhetorical tendencies, forming a mid-level abstraction of the community's communicative identity:

$$D_c = \mathbf{LLM}(\mathcal{A}_c^{\text{top-}K}, P_d) \tag{4}$$

History Persona. To endow community c with memory, we retrieve semantically aligned past experiences in response to a new event e_k. We encode the event and each post using instruction-tuned embeddings that prioritize intent, narrative frame, and ideological context:

$$\begin{aligned} X_{e_k} &= \mathbf{LLM}_{embed}(e_k, P_k) \\ X_{a_j} &= \mathbf{LLM}_{embed}(a_j) \end{aligned} \tag{5}$$

The top-M most relevant posts are selected via cosine similarity:

$$\Psi(X_{e_k}, X_{a_j}) = \frac{X_{e_k} \cdot X_{a_j}}{\|X_{e_k}\| \|X_{a_j}\|} \tag{6}$$

These retrieved posts $\mathcal{N}_c^{\text{top-}M}$ form the *history persona*, reflecting the community's contextual precedents.

Influencer Endorsement. To capture intra-community narrative anchoring, we simulate endorsements from top-H influencers in community c, selected by follower count. Given an influencer u_h's recent posts $\mathcal{I}_{u_h}$, the LLM is prompted to generate their reaction to e_k:

$$r_{u_h}^{(k)} = \mathbf{LLM}(\mathcal{I}_{u_h}, e_k, P_i) \tag{7}$$

The aggregated community-level influencer response is:

$$R_{c,k}^{\text{top-}H} = \frac{1}{H} \sum_{h=1}^{H} R_{u_h}^{(k)} \tag{8}$$

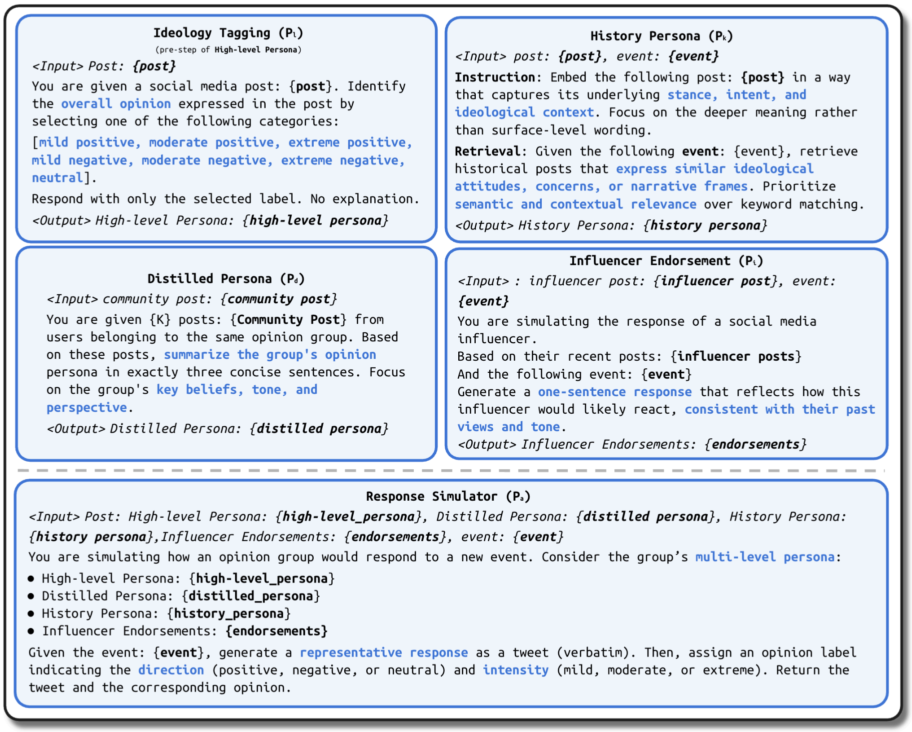


Fig. 4: Prompt Interface for Simulating Ideological Community Responses.

Simulated Response. Combining the symbolic (H_c), narrative (D_c), experiential ($\mathcal{N}_c^{\text{top-}M}$), and influencer-driven ($R_{c,k}^{\text{top-}H}$) perspectives, we simulate how community c would collectively respond to e_k:

$$R_{a_c^k}, s_{a_c^k} = \textbf{LLM}(H_c, D_c, \mathcal{N}_c^{\text{top-}M}, r_{c,k}^{\text{top-}H}, e_k, P_a) \tag{9}$$

Here, $R_{a_c^k}$ is a representative tweet and $s_{a_c^k}$ the predicted stance (direction and intensity).

User Opinion Update. Following the community response, each user u_i belonging to community c updates their stance via inertia-weighted assimilation:

$$s_{u_i}^{(t)} = \frac{\lambda}{\eta_i} s_{u_i}^{(t)} + \left(1 - \frac{\lambda}{\eta_i}\right) s_{a_c^k} \tag{10}$$

where η_i denotes the user's consistency coefficient and λ is a global resistance parameter. This formulation reflects classic social influence dynamics, where individuals assimilate toward group norms in proportion to their ideological malleability. The full set of prompts used for community persona construction is shown in Figure 4.

4.3 Dynamic Ideology Simulation

We simulate evolving user stances through a structured process of neighbor selection, message aggregation, and opinion update (shown in the *Dynamic Ideology Simulation* of Figure 2), inspired by classical opinion dynamics theories and illustrated in [6].

Neighbor Selection. To identify potential influencers at each time step, user u_i selects a dynamic neighbor set $J_{u_i}^{(t)}$ incorporating both social connections and opinion alignment. Guided by social identity and bounded confidence theory [12, 7], two neighborhoods are defined:

$$\begin{aligned} J_{u_i,t}^{\text{opinion}} &= j \mid |s_i - s_j| \leq \epsilon, \quad \epsilon = \frac{\alpha}{\eta_i} \\ J_{u_i,t}^{\text{network}} &= j \mid (i,j) \in \mathcal{E} \\ J_{u_i}^{(t)} &= J_{u_i,t}^{\text{opinion}} \cup J_{u_i,t}^{\text{network}} \end{aligned} \tag{11}$$

where α is a global confidence threshold and η_i encodes user u_i's resistance to ideological deviation—users with higher η exhibit narrower acceptable opinion bands.

Message Aggregation. Each user emits a message $m_{u_i}^{(t)}$ that reflects their stance. Unless specified otherwise, we assume transparent messaging:

$$m_{u_i}^{(t)} = f_{\text{message}}(s_{u_i}^{(t)}) = s_{u_i}^{(t)} \tag{12}$$

This reflects a widely-used assumption in ABMs that agents truthfully communicate their internal beliefs [6].

Opinion Update. User opinions evolve in response to messages from selected peers and external stimuli. OpinioNet supports three canonical models:

Friedkin-Johnsen (FJ) [12]: Updates balance prior conviction and peer influence, modulated by ideological inertia:

$$s_i^{(t+1)} = \frac{\lambda}{\eta_i} s_i^{(0)} + \left(1 - \frac{\lambda}{\eta_i}\right) \cdot \frac{1}{|J_{u_i}^{(t)}|} \sum_{j \in J_{u_i}^{(t)}} s_j^{(t)} \tag{13}$$

Hegselmann-Krause (HK) [19]: Agents update by averaging over the full set of compatible neighbors:

$$s_i^{(t+1)} = \frac{1}{|J_{u_i}^{(t)} \cup i|} \sum_{j \in J_{u_i}^{(t)} \cup i} s_j^{(t)} \tag{14}$$

Deffuant-Weisbuch (DW) [8]: Models pairwise convergence under bounded confidence. We randomly sample $j \in J_{u_i}^{(t)}$:

$$\begin{aligned} &\text{if } |s_i^{(t)} - s_j^{(t)}| \leq \epsilon : \\ &s_i^{(t+1)} = s_i^{(t)} + \mu(s_j^{(t)} - s_i^{(t)}) \\ &s_j^{(t+1)} = s_j^{(t)} + \mu(s_i^{(t)} - s_j^{(t)}) \end{aligned} \tag{15}$$

where $\mu \in (0, 0.5]$ is the convergence rate.

Table 1: Statistics of the Twitter and GDELT datasets used in our experiments. The news events are sourced from GDELT and filtered by dataset-specific keywords and event location.

Dataset	Time Range	# Users	# Posts (Train / Test)	# Claims (Train / Test)	GDELT News Filter (Keyword / Location)
Russia Slogans	Aug 13 – Oct 31, 2022	274	3911 / 3094	1444 / 1560	*Russia, Ukraine, NATO* / US
Russophobia	Aug 04 – Aug 25, 2022	2377	9316 / 8296	4119 / 4054	*Russophobia, NATO* / US
EDCA	Jul 04 – Aug 25, 2022	5204	5858 / 9154	4593 / 3840	*EDCA, Philippines, US Military* / US and Philippines

Table 2: Performance (%) of OpinionNet under Various Social Influence Models. Δ Represents the Average Improvement Over the Corresponding Pure Baseline Across Datasets. The Maximum Value Across All Models is Bolded.

Method	Russia Slogans			Russophobia			EDCA			Δ (↑)		
	F1	M-F1	r	F1	M-F1	r	F1	M-F1	r	F1	M-F1	r
Pure FJ	64.26	60.17	22.70	51.30	48.62	38.54	68.40	62.37	50.87	–	–	–
+OpinionNet	84.41	79.35	37.62	52.74	48.68	48.66	72.17	67.74	62.31	**13.79**	**14.38**	**32.54**
w/o H_c	80.35	75.47	34.16	52.01	48.76	45.53	71.37	64.28	59.81	10.76	10.14	24.43
w/o D_c	76.39	72.57	34.23	51.69	48.67	41.94	70.29	62.86	55.93	7.99	7.56	17.83
w/o $\mathcal{N}_c^{top\text{-}M}$	78.84	72.93	33.89	51.57	48.66	43.27	71.89	63.68	**67.74**	10.13	8.24	29.25
w/o $\mathcal{R}_{c,k}^{top\text{-}H}$	77.54	74.23	34.51	51.93	48.57	42.85	71.02	63.06	57.88	9.15	8.59	20.63
Pure HK	56.65	52.69	30.42	50.17	48.68	61.83	67.32	63.31	51.21	–	–	–
+OpinionNet	81.37	76.74	**53.41**	**61.73**	**57.26**	**70.66**	66.79	67.74	59.22	**20.53**	**22.50**	**27.76**
w/o H_c	78.52	74.21	53.07	58.84	54.46	67.40	67.28	66.84	56.66	17.51	18.88	23.47
w/o D_c	74.61	71.08	47.46	59.83	54.25	65.27	68.02	66.68	57.52	16.26	16.75	18.67
w/o $\mathcal{N}_c^{top\text{-}M}$	76.92	72.27	50.84	57.67	56.29	66.68	66.26	67.94	57.24	15.34	19.48	21.82
w/o $\mathcal{R}_{c,k}^{top\text{-}H}$	77.64	74.27	45.78	57.02	48.53	66.42	67.94	67.07	58.14	16.34	15.45	18.74
Pure DW	55.13	52.74	28.48	51.37	47.91	23.15	70.55	66.42	47.54	–	–	–
+OpinionNet	**86.69**	**82.36**	50.95	53.70	49.02	46.27	**73.07**	**69.67**	60.49	**20.56**	**20.75**	**59.03**
w/o H_c	85.78	80.94	50.37	52.94	47.75	45.23	72.68	67.96	59.83	19.40	17.71	56.73
w/o D_c	81.69	73.52	42.78	52.19	48.70	39.48	71.76	68.23	52.28	16.18	13.99	35.67
w/o $\mathcal{N}_c^{top\text{-}M}$	83.73	78.52	48.96	52.08	48.94	42.49	70.38	69.35	55.16	16.46	17.80	47.84
w/o $\mathcal{R}_{c,k}^{top\text{-}H}$	82.98	76.40	48.72	51.67	48.12	43.31	70.08	68.38	54.95	15.63	15.46	48.21

Together, these mechanisms simulate how users adjust opinions in response to peers, group norms, and news stimuli—anchored by personal ideological rigidity η_i.

5 Experiment

5.1 Experiment Setting

In our experiments, the backbone large language model (LLM) used in OpinioNet is GPT-3.5. The following hyperparameters are used throughout: $\alpha = 1$, $\lambda = 0.5$, and $\mu = 0.5$. Each simulation spans 20 discrete time steps to capture evolving stance dynamics.

To simulate community-level responses, we incorporate three types of personas for each opinion group. The **Distilled Persona** is constructed by selecting the top $K = 20$ representative posts from the training set of the community. The **History Persona** for each user is retrieved using the top $M = 10$ most semantically similar historical posts based on embedding similarity. Additionally, we identify $H = 5$ influencer users per opinion community using historical engagement frequency, and include their most popular assertions during external events to guide community stances.

5.2 Datasets

Table 1 summarizes the statistics of the datasets used in our experiments. Each dataset captures a distinct geopolitical discourse on Twitter, with varying temporal spans and ideological focuses. To ensure meaningful stance evolution modeling, we retain users who contribute sufficiently in both training and testing phases. For external event simulation, we collect news articles from the GDELT Global Knowledge Graph (GKG), filtered by topic-specific keywords and event location.

For the **Russia Slogans** and **Russophobia** datasets, we focus on events reported within the U.S. that are related to Russia, Ukraine, and NATO. For the **EDCA** dataset, which centers on U.S.–Philippines defense cooperation, we include events reported from both the U.S. and the Philippines to capture region-specific narratives. The curated news events serve as realistic external shocks for simulating ideology-aware opinion dynamics across polarized user communities.

5.3 Task Formulation and Evaluation Metrics

We evaluate OpinioNet across two key tasks: User Opinion Prediction and Opinion Dynamics Trend Prediction, each designed to assess the system's ability to simulate both discrete and continuous ideological behavior.

User Opinion Prediction. Given a time series of predicted stance scores ${s_{u_i}^{(t)}}_{t=1}^{T}$ for user u_i, we compute the user's final opinion as the sign of their average stance:

$$\mathrm{sgn}\left(\frac{1}{T}\sum_{t=1}^{T} s_{u_i}^{(t)}\right) \tag{16}$$

To evaluate this classification task, we adopt:

- **F1 (Micro-F1)**: The harmonic mean of precision and recall across all users, favoring performance on dominant groups.
- **M-F1 (Macro-F1)**: The unweighted average F1 across all classes, emphasizing fair treatment of both majority and minority opinions.

Opinion Dynamics Trend Prediction. We further assess how well OpinioNet simulates the aggregate ideological trend over time. The average opinion trajectory across all N users is computed as:

$$\bar{s}^{(t)} = \frac{1}{N}\sum_{i=1}^{N} s_{u_i}^{(t)}, \quad H = [\bar{s}^0, \bar{s}^1, \ldots, \bar{s}^T] \tag{17}$$

To evaluate this continuous prediction task, we use:

- r **(Pearson Correlation)**: A standard measure of linear association between the predicted and true opinion trajectories, capturing the fidelity of temporal opinion trends [20].

Together, these metrics provide a holistic evaluation of both user-level stance classification and population-level ideological simulation.

5.4 Baselines

To benchmark the performance of OpinioNet, we compare it against classical Agent-Based Models (ABMs) that simulate both users and news sources as interacting agents. In these baselines, news domains are treated equivalently to user agents, and opinion updates follow the dynamics outlined in Section 4.3. We implement the following widely-used ABM frameworks: (1) the Friedkin–Johnsen (FJ) model [11], (2) the Hegselmann–Krause (HK) model [19], and (3) the Deffuant–Weisbuch (DW) model [8].

To ensure a fair comparison, we apply a unified set of hyperparameters across all baselines and OpinioNet: $\alpha = 1$, $\lambda = 0.5$, and $\mu = 0.5$.

5.5 Main Result

User Opinion Prediction. Table 2 shows that OpinioNet consistently and substantially improves over traditional Agent-Based Models (FJ, HK, DW) across all datasets in terms of both Micro-F1 and Macro-F1. For instance, when applied on top of the FJ model, OpinioNet boosts Macro-F1 by an average of 14.38%, confirming its ability to better capture minority and balanced group opinions. This improvement is even more pronounced when integrated with the DW model, achieving the highest F1 (86.69%) and Macro-F1 (82.36%).

These gains highlight OpinioNet's superior capability in modeling user-level stance shifts, driven by its use of high-level and distilled personas to simulate ideological behavior in response to external signals. The model benefits from structured social memory and event-aware group simulation, which help encode ideological bias and historical behavior more effectively than simple numerical opinion updates in classical models.

Opinion Dynamics Trend Prediction. In predicting the temporal trend of average user opinions, OpinioNet again outperforms all ABM baselines with large margins in Pearson correlation (r). Particularly under the DW framework, it improves r by +59.03%, indicating much stronger alignment with real-world aggregate opinion trajectories. This suggests that OpinioNet is able to capture not only accurate discrete user stances but also fine-grained temporal shifts in community sentiment, a capability critical for social forecasting.

Ablation Results. Removing key components of OpinioNet leads to noticeable performance drops across all tasks. Excluding the high-level persona (H_c) consistently reduces performance, supporting its role in representing ideology-guided discourse. Similarly, removing the distilled persona (D_c) hurts both F1 and correlation, especially in DW (14.71% drop in r), reflecting its importance in capturing group-specific linguistic style. Eliminating top-M neighbors ($\mathcal{N}_c^{top\text{-}M}$) or top-$H$ endorsements ($\mathcal{R}_{c,k}^{top\text{-}H}$) also degrades performance, underscoring the value of social influence and endorsement modeling in opinion dynamics.

Qualitative Case Study: Simulated Community Responses to EDCA News. To illustrate the ideological sensitivity and contextual realism of OpinioNet's simulation process, we present a qualitative case study centered on Exercise KAMANDAG 2022—an event conducted under the U.S.–Philippines Enhanced Defense Cooperation Agreement

Table 3: Simulated Community Responses to EDCA News

Ideological Community	Simulated Response
Extreme Positive	"Joint exercises like KAMANDAG are vital for securing peace and promoting regional stability. The U.S.-Philippines alliance stands as a beacon of democracy and humanitarian cooperation in the Indo-Pacific."
Neutral	"While joint military operations enhance disaster response and coordination, regional actors must tread carefully to avoid fueling unnecessary military competition."
Extreme Negative	"KAMANDAG is yet another show of U.S. imperialist aggression under the guise of 'security.' These drills threaten sovereignty and escalate militarization in Asia."

(EDCA). In this joint operation, over 3,000 U.S. and Philippine Marines, along with troops from Japan and South Korea, participated in amphibious training and humanitarian readiness exercises.

We prompted ideological community agents to respond to a curated news summary of this event. Table 3 presents exemplar responses generated from three ideological clusters: **Extreme Positive**, **Neutral**, and **Extreme Negative**. These responses reflect distinct rhetorical strategies grounded in each community's symbolic stance, historical post memory, and influencer endorsements.

– The **Extreme Positive** community celebrated the exercise as a "beacon of democracy," emphasizing alliance strength and humanitarian benefits.
– The **Neutral** group offered a balanced perspective, acknowledging operational gains while warning against "regional military competition."
– The **Extreme Negative** group framed the event as "U.S. imperialist aggression," invoking themes of sovereignty and militarization.

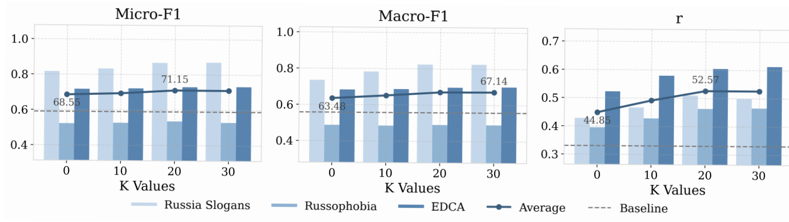
These results highlight OpinioNet's ability to generate linguistically coherent, ideologically aligned narratives. The simulation respects not only the external event content but also the nuanced belief systems encoded in each community's multi-level persona, demonstrating the framework's capacity to model real-world discourse in politically sensitive contexts.

Summary. Overall, OpinioNet delivers robust, interpretable, and impactful improvements across all baselines and datasets. Its architecture—combining language-aware community personas and selective influencer modeling—offers a powerful new paradigm for simulating opinion evolution beyond what numeric opinion exchange models can achieve.

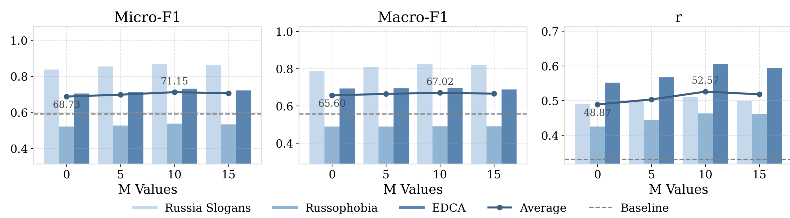
5.6 Sensitivity Analysis

To investigate the impact of key parameters and validate the design choices in OpinioNet, we conduct sensitivity analyses on three critical components: the number of posts used to construct the distilled persona (K), the number of retrieved historical posts for user history (M), and the number of influencers used for endorsement modeling (H). The results are summarized in Figure 5.

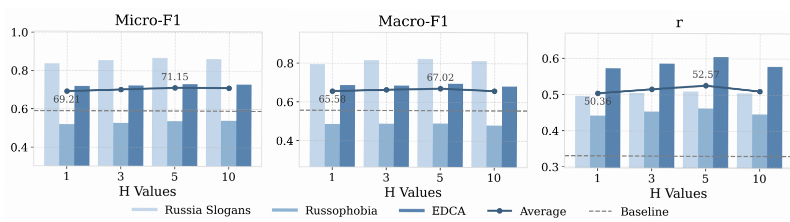
Effect of K on Distilled Persona (D_c). As shown in the second row of Figure 5a, increasing K (number of posts used to generate the distilled persona) consistently



(a) Model performance across different K used in Distilled Persona (D_c).



(b) Model performance across different M used in History Persona ($\mathcal{N}_c^{top-M}$).



(c) Model performance across different H used in Influencer Endorsement ($\mathcal{R}_{c,k}^{top\text{-}H}$).

Fig. 5: Sensitivity Analysis of OpinioNet: Model performance under varying values of three key parameters. The average line is annotated with the maximum and minimum points.

improves performance across all metrics—F1, M-F1, and r. This trend validates the benefit of incorporating more representative ideological content into community personas. However, the marginal gain diminishes beyond $K = 20$, likely due to token budget saturation and diminishing informativeness per additional post. Therefore, we adopt $K = 20$ as the default setting, balancing performance and computational cost.

Effect of M on History Persona ($\mathcal{N}_c^{top-M}$). We observe a similar pattern in the third row of Figure 5b when varying M, which controls the number of retrieved posts for constructing individual history personas. Performance peaks at $M = 10$, after which excessive historical context may introduce outdated or noisy signals. This aligns with our intuition that recent, highly relevant posts are most informative for capturing a user's current ideological leaning. We thus set $M = 10$ to maximize informativeness while avoiding overfitting to historical noise.

Effect of H on Influencer Endorsement ($\mathcal{R}_{c,k}^{top\text{-}H}$). The first row in Figure 5c shows the effect of varying H, the number of high-influence users whose posts guide community responses. Increasing H from 1 to 5 significantly improves performance, particularly in Micro-F1 and correlation, suggesting that a small set of representative influencers can effectively guide ideological response simulation. Performance slightly plateaus beyond $H = 5$, indicating that additional influencers may introduce heterogeneity or conflicting stances. We set $H = 5$ to achieve ideological clarity without sacrificing generalizability.

Implications and Design Takeaways. This analysis supports OpinioNet's modular design: each persona component—distilled (D_c), history-based ($\mathcal{N}_g^{top-M}$), and influencer-based ($\mathcal{R}_{c,k}^{top\text{-}H}$)—plays a unique role in driving predictive accuracy. The selected defaults ($K = 20$, $M = 10$, $H = 5$) strike a robust balance between expressiveness and noise control, facilitating both accurate user-level predictions and realistic macro-level opinion trend modeling. The consistent improvements across diverse metrics and datasets highlight the adaptability and generalizability of OpinioNet across sociopolitical contexts.

6 Conclusion

We present OpinioNet, a novel LLM-driven framework for simulating opinion dynamics in ideologically polarized communities. By integrating multi-level personas, event-triggered simulation, and social influence modeling, OpinioNet enables a fine-grained, language-aware understanding of how opinions evolve over time. Our extensive experiments across real-world geopolitical datasets demonstrate that OpinioNet consistently outperforms traditional agent-based models in both classification accuracy and dynamic trend prediction. Beyond performance, OpinioNet introduces a scalable and interpretable mechanism for incorporating external stimuli, such as news events, into user-level stance shifts. This work marks a significant step toward bridging the gap between symbolic social theories and neural simulation, offering a powerful tool for studying ideological polarization, misinformation response, and public opinion forecasting in high-stakes sociopolitical contexts.

7 Limitations

While OpinioNet demonstrates strong performance in simulating opinion dynamics, several limitations remain. First, the framework relies on predefined community clusters, which may oversimplify nuanced ideological spectra and overlook emergent user realignments. Second, the simulation quality depends on the fidelity of LLM-generated responses, which may reflect biases in the underlying model or training data. Third, although we incorporate real-world events via GDELT, our selection is restricted to a finite set of media sources and keywords, potentially limiting coverage and granularity. Addressing these challenges—such as dynamically evolving community structures, bias-controlled LLM reasoning, and event generalization—presents important directions for future work.

8 Acknowledgments

Research reported in this paper was sponsored in part by the DARPA award HR001121C0165, the DARPA award HR00112290105, the DoD Basic Research Office award HQ00342110002, the Army Research Laboratory under Coop- erative Agreement W911NF-17-20196.

References

1. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al.: Gpt-4 technical report. arXiv preprint arXiv:2303.08774 (2023)
2. Aher, G., Arriaga, R.I., Kalai, A.T.: Using large language models to simulate multiple humans. arXiv preprint arXiv:2208.10264 **5** (2022)
3. Argyle, L.P., Busby, E.C., Fulda, N., Gubler, J.R., Rytting, C., Wingate, D.: Out of one, many: Using language models to simulate human samples. Political Analysis **31**(3), 337–351 (2023)
4. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al.: Language models are few-shot learners. Advances in neural information processing systems **33**, 1877–1901 (2020)
5. Centola, D., Macy, M.: Complex contagions and the weakness of long ties. American journal of Sociology **113**(3), 702–734 (2007)
6. Chuang, Y.S., Rogers, T.T.: Computational agent-based models in opinion dynamics: A survey on social simulations and empirical studies. arXiv preprint arXiv:2306.03446 (2023)
7. Dandekar, P., Goel, A., Lee, D.T.: Biased assimilation, homophily, and the dynamics of polarization. Proceedings of the National Academy of Sciences **110**(15), 5791–5796 (2013)
8. Deffuant, G., Neau, D., Amblard, F., Weisbuch, G.: Mixing beliefs among interacting agents. Advances in Complex Systems **3**(01n04), 87–98 (2000)
9. van Dijk, T.A.: Ideology: A Multidisciplinary Approach. Sage (1998)
10. Eagly, A.H., Chaiken, S.: The Psychology of Attitudes. Harcourt Brace (1993)
11. Friedkin, N.E., Johnsen, E.C.: Social influence and opinions. The Journal of Mathematical Sociology **15**(3-4), 193–206 (1990). https://doi.org/10.1080/0022250X.1990.9990069, https://doi.org/10.1080/0022250X.1990.9990069
12. Friedkin, N.E., Johnsen, E.C.: Social influence networks and opinion change. Advances in group processes **16**(1), 1–29 (1999)
13. Gao, C., Lan, X., Li, N., Yuan, Y., Ding, J., Zhou, Z., Xu, F., Li, Y.: Large language models empowered agent-based modeling and simulation: A survey and perspectives. Humanities and Social Sciences Communications **11**(1), 1–24 (2024)
14. Latoski, L.C.F., Dantas, W., Arenzon, J.J.: Opinion inertia and coarsening in the persistent voter model. Physical Review E **109**(5), 054115 (2024)
15. Martins, A.C.: Continuous opinions and discrete actions in opinion dynamics problems. International Journal of Modern Physics C **19**(04), 617–624 (2008)
16. Moussaïd, M., Kämmer, J.E., Analytis, P.P., Neth, H.: Social influence and the collective dynamics of opinion formation. PloS one **8**(11), e78433 (2013)
17. Park, J.S., O'Brien, J., Cai, C.J., Morris, M.R., Liang, P., Bernstein, M.S.: Generative agents: Interactive simulacra of human behavior. In: Proceedings of the 36th annual acm symposium on user interface software and technology. pp. 1–22 (2023)
18. Petty, R.E., Krosnick, J.A.: Attitude strength: Antecedents and consequences. Psychology Press (2014)
19. Rainer, H., Krause, U.: Opinion dynamics and bounded confidence: models, analysis and simulation (2002)
20. Sedgwick, P.: Pearson's correlation coefficient. Bmj **345** (2012)
21. Tajfel, H., Turner, J.C.: An integrative theory of intergroup conflict. In: The social psychology of intergroup relations, pp. 33–47. Brooks/Cole (1979)
22. Zino, L., Giardini, F., Vilone, D., Cao, M.: On modeling collective risk perception via opinion dynamics. European Journal of Control **80**, 101036 (2024)