A parameter-free model for the online spread of far-right messages: combining Agent-Based Models with Large-Language Models

Stephen Zhong[1][0009−0001−2688−6108], Nathalie Japkowicz[1][0000−0003−1176−1617], Frédéric Amblard[2][0000−0002−2653−0857], and Philippe J. Giabbanelli[3][0000−0001−6816−355X]

[1] Dept. of Computer Science, American University, Washington, DC 20016, USA {sz8367a,japkowic}@american.edu
[2] IRIT, Universite de Toulouse, France frederic.amblard@ut-capitole.fr
[3] Virginia Modeling, Analysis, and Simulation Center (VMASC), Old Dominion University, Suffolk, VA 23435, USA pgiabban@odu.edu

Abstract. Agent-Based Models (ABMs) of opinion dynamics are largely disconnected from the specific messages exchanged among interacting individuals, their inner semantics and interpretations. Rather, ABMs often abstract this aspect through corresponding numerical values (e.g., -1 as against and +1 as totally in favor). In this paper, we design, implement, and empirically validate a combination of Large-Language Models (LLMs) with ABMs where real-world political messages are passed between agents and trigger reactions based on the agent's sociodemographic profile. Our computational experiments combine real-world social network structures, posting frequencies, and extreme-right messages with nationally representative demographics for the U.S. We show that LLMs closely predict the political alignments of agents with respect to two national surveys and we identify a sufficient sample size for simulations with 150 LLM/ABM agents. Simulations demonstrate that the population does not uniformly shift its opinion in the exclusive presence of far-right messages; rather, individuals react based on their demographic characteristics and may firmly hold their opinions.

Keywords: Belief spread · Hybrid Model · Online social network.

1 Introduction

The rise of extreme right ideologies on online social media platforms, such as X (formerly Twitter), has become an important phenomenon with profound social and political implications. These ideologies, often characterized by hate speech, misinformation, and polarizing rhetoric, have found fertile ground in the digital age. Algorithms amplify the natural tendency of individuals to prefer more extreme views within their political group [63], particularly by amplifying the political right [27]. Amplification contributes to creating echo chambers that foster radicalization [43], even if the concept of an echo chamber is approached

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_16

differently across studies [35]. Characterizing the dynamics for the spread of far-right online messages is critical, as the influence of 'e-extremism' [61] extends beyond the virtual realm, contributing to real-world violence [9], the erosion of democratic norms, and the marginalization of vulnerable communities [56].

Modeling the spread of beliefs in online social networks continues to be a fertile area of research, as exemplified by multiple empirical studies at the International Conference on Computational Science [32, 21]. Models specialized in the spread of hate speech need to account for several characteristics. First, although the far-right may share some narratives (e.g., collective victimhood), it is composed of extremely heterogeneous organizations [18] and individuals with different motives. Online hate thus varies substantially across users [5]. Given this heterogeneity, several models explicitly represent each individual instead of grouping them into aggregates assumed to behave identically. A commonly used technique is Agent-Based Modeling (ABM) has been particularly used, where each entity has its own attributes and/or rules and interacts with others in a local environment that may be digital or physical [39, 11, 52]. Second, ideas do not exactly spread like viruses: instead, there is a *gradual* build up in a person's beliefs and attitudes [46]. From a simulation standpoint, models thus often track extremism among individual agents using a numerical scale rather than through categorical states (e.g., susceptible or 'infected' with extreme ideas). Although complex ABMs may not allow us to identify the analytical solutions afforded by simpler mass action models of political extremism [12], they are helpful in identifying tailored solutions [58, 57] (e.g., for different user profiles, behaviors, locations), improve accuracy by incorporating spatial and network effects [48], and they can estimate uncertainty by capturing stochasticity at the individual level. Despite these advantages, current ABMs for the spread of far-right messages have two important limitations, summarized as follows.

First, in current models, *agents do not exchange actual texts*; rather, their interactions are abstracted as a stochastic process such as the probability of passing a type of message, or an agent gradually aligns itself on the state most commonly encountered among its peers [42]. A model is a simplification, and this longstanding abstraction of text evades the complexity of text processing while answering important questions. But without the text, we miss an important marker for detecting and preventing violence [15]. Furthermore, without knowing how agents react to specific messages, we cannot estimate the effect of campaigns to debunk specific arguments, such as COVID or election conspiracies [36].

Second, several parameters were created to keep the ABMs simple, such as the 'ease' or 'volatility' at which agents would change opinions [11], the strength at which they would 'influence' others [8], or their 'tolerance' threshold to other opinions [59]. These are called *free parameters* [30] as they *are very difficult to calibrate empirically* [11, 59, 8] since they do not directly map to a real-world characteristic. Their combined values are calibrated by comparing aggregate outcomes for the overall ABM with expectations, but their individual values cannot be known [30]. For example, there does not exist a general 'ease' at which somebody changes their mind: it depends on the person and the message, among

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_16

other aspects. Other fields have stressed the importance of empirical grounding for ABMs of social spreads. For instance, a review on innovation diffusion emphasized that the ability to calibrate ABMs from data is instrumental to shift their use from a learning tool onto guiding policy decisions [60]. We thus need ABMs with minimal reliance on parameters that cannot be individually calibrated.

Our main contribution is to address both limitations by avoiding the use of parameters and by supporting the spread of actual messages. This is achieved by combining ABMs with Large-Language Models (LLMs), such that each agent uses a LLM to model its reactions with respect to a specific message based on the agent's key demographics (age, sex, race and ethnicity, educational attainments). In this paper, we design, implement, and assess the hybrid ABM/LLM model on a sample case study using empirical data that includes U.S. demographics, political opinions, and messages from the Truth Social platform.

The remainder of this paper is organized as follows. In Section 2, we succinctly cover the design of ABMs for the spread of far-right messages or ideas that marginalize vulnerable social groups, along with the emerging practice of hybrid ABM/LLM models for computational social science. In Section 3, we present the design of our model and explain its empirical grounding in representative data sources. Experiments in Section 4 show how agents change in reaction to different messages, at several population scales. Finally, we summarize the core limitations of the present model in Section 5 and provide directions for extensions. To support replicability, our model and experiments are accessible on a third-party repository at https://osf.io/h8zme/.

2 Background

2.1 Design of ABMs for the spread of far-right ideas

Building upon foundational work in opinion dynamics from the 1960s [1] and 1970s [14], the past two decades have witnessed a surge in ABM applications to this field, exemplified by seminal contributions such as those by Deffuant *et al.* [13] and Hegselmann & Krause [47], among many others (see [33] for a comprehensive overview). This rich literature reveals key distinctions among models, which we organize in the five following critical characteristics.

First, most models represent opinions numerically (binary, discrete, or continuous) facilitating straightforward measurement of opinion distances and thus, the quantification of influence processes among agents. Interestingly, this numerical encoding parallels methods used in political science to represent political actors within multidimensional ideological spaces [54, 31].

Second, the number of actors involved in each influence event also varies. While many models focus on peer-to-peer interactions, involving only two agents, other configurations have been investigated. Many-to-one influence, where a single individual is influenced by multiple agents (e.g., averaging opinions in the immediate social environment [47]), and one-to-many influence, better suited for information diffusion [51], are notable examples.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_16

Third, the core process of influence itself forms another significant point of divergence among ABM models. For a considerable period, mimetic influence– the tendency for agents to adopt opinions similar to those observed in their social environment– dominated the literature [16]. However, the inclusion of contrarian dynamics [19] allows to model behaviors such as radicalization or the deliberate distancing of opinions. More sophisticated models integrate both mimetic and contrarian dynamics, making the influence process dependent on the opinion distance between interacting individuals. Agents with similar views converge, while those with dissimilar views diverge even further [28].

Fourth, the *substratum* of social influence–the underlying social structure within which interactions occur– represents another key variable. Early models often employed assumptions of random mixing within the population. More recent work range from abstract network models (e.g., small-world or scale-free networks) to the incorporation of empirically derived social network data [10], often producing more nuanced and realistic results.

Finally, the lack of longitudinal data on individual opinions, often due to sensitivity concerns, limits the evaluation of accuracy and predictive power of ABMs. While macroscopic-level snapshots of opinion distributions are available, the absence of detailed individual-level data restricts comprehensive validation efforts. Experimental data from social psychology and ethology provide partial micro-validation of specific influence mechanisms, but often fall short of capturing complex real-world influence processes, such as the role of media [16].

2.2 Combining ABMs with Large Language Models

Combining ABMs with LLMs creates a *hybrid* (systems) model since it uses a simulation technique along with a technique from another domain [40]. As this specific type of hybrid is relatively new, it goes by different names such as 'LLM-based agents' [20] or 'Generative Agent-Based Modeling' [23]. Several works have either proposed or demonstrated that describing the sociodemographic characteristics [45, 4] of agents (i.e., 'conditioning' a prompt or creating a 'persona') can then "leverage the vast data within LLMs to capture human behavior and decision-making [instead of] relying on modelers' assumptions" [23].

A conceptual framework for disinformation research and LLMs proposed to power agents within a social network via LLMs [45]. GPT-4 was viewed as a promising tool to suggest the evolution of opinions given a user profile and exposure to an idea such as electoral fraud. Several of these ideas were realized in a study by Zheng and Tang, released in November 2024 [62], who created a small model where agents interact on Twitter (post, retweet, reply, like). Operating without empirical data, the model illustrated changes in attitudes on an abstraction of the Roe v. Wade case on abortion. Although no statistical analyses were conducted, visualizations suggested that average attitudes may fluctuate over time without ever stabilizing, depending on the synthetic network topology employed (small-world vs. scale-free networks). Opinion diversity (measured as the number of unique opinions) also depended on the topology by increasing in one case (small-world) and decreasing in the other (scale-free) [62].

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_16

3 Methods

3.1 Design of the hybrid ABM/LLM model

Overview. We initialize each agent in our population with a set of demographic characteristics (age, sex, race and ethnicity, educational attainments) that are partially predictive of their initial opinion score. Agents are connected to mimic the follower relationships on mainstream social media such as Twitter (before becoming X). A set of right-leaning agents post specific messages based on an empirical frequency that accounts for the relation between the amount of posts and the number of followers (i.e., agents with fewer followers post less). In this one-to-many influence model (section 2.1), all followers of the posting agents will read their posts. Followers react immediately upon reading by using a LLM that accounts for the content of the messages and the reader's demographic characteristics The LLM is tasked with *suggesting* a new plausible opinion score, which may become more conservative or more liberal as the LLM integrates contrarian and mimetic dynamics (section 2.1). Since humans do not widely change opinions by reading a single post, the suggested score is compared with the agent's current opinion and leads to a moderate update that follows the empirical literature on gradual changes in opinions. Previous models have shown that dynamics may embrace a chaotic or an oscillatory trajectory [62], thus we end the simulation after a set number of steps rather than stabilization.

Formal description. The population is modeled as a graph $G = (V, E)$ consisting of a set of users V and the users whom they follow via directed edges E. Each agent $i \in V$ has a constant set of demographic characteristics i_{dem} over the duration of the simulation, and a variable opinion score $i^t_{opi} \in [-10, 10]$ that is updated over discrete time ticks t. At $t = 0$, we use a LLM to initialize the scores based on each agent's demographics, that is, $i^0_{opi} \leftarrow LLM(i_{dem}) \forall i \in V$. Positive scores indicate left-leaning agents and negative scores indicate right-leaning agents. The simulation proceeds for a duration specified by the user.

At each tick $t > 0$, we perform an *asynchronous update* in three steps. First, right-leaning agents $\{i | i \in V, i^t_{opi} > 0\}$ have a probability $P(i) = f(d_{in}(i))$ of posting based on their number of followers, that is, the number of incoming edges $d_{in}(i) = |\{u \in V | (u, i) \in E\}|$. A specific message is chosen at random from a set of far-right messages $\mathbb{M}$. Second, for all agents who post, their followers $\{j | i, j \in V, i^t_{opi} > 0, (j, i) \in E\}$ read and react to the messages. That is, the LLM is tasked with *suggesting* a *plausible numerical* score based on the reader's demographics and the message content, $LLM(j_{dem}, m \in \mathbb{M})$. Due to stochastic variations in the LLM, it may not deliver a numerical score, it may not be within the target range, or it may be an implausibly wide departure from the agent's current opinion. We thus treat $LLM(j_{dem}, m \in \mathbb{M})$ as a random variable F that starts by drawing a sample F' and calls itself again if criteria are not yet met:

$$\begin{cases} F'(j_{dem}, m) & \text{if } F'(j_{dem}, m) \in [-10, 10] \text{ and } |j^t_{\text{opi}} - F'(j_{\text{dem}}, m)| \leq 2, \\ F(j_{dem}, m) & \text{otherwise} \end{cases} \quad (1)$$

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_16

Unlike in a synchronous update where agents' values are buffered until all have been visited, the asynchronous update computes and updates values in the order in which agents are visited. This order may change at each time tick, as agents are updated in reaction to a stochastic event (following an account that posted). This mechanism thus uses an asynchronous update with random order.

Implementation considerations. Engineering the prompt is an essential component of a LLM-based system. The prompt used to update an agent is shown in Box 1. We experimented several versions of this wording. Suggesting that the post was sent *by a friend* or *by a peer* had a risk of biasing the LLM in trusting the message, thus we removed any description of the sender. This illustrates the well-documented notion that less can be better in a prompt [37]. We had to engage in 'roleplay' by including "*pretend you are*" in the prompt, otherwise the LLM would state that it is a machine learning model that cannot give an opinion. As it is stochastic (even with a temperature of 0), the LLM may occasionally return an invalid response, such as a variation of "I cannot answer this" that lacks an opinion score. We thus loop queries to the LLM until a satisfactory answer is obtained, in line with other recent works on prompting [53].

Box 1. Prompt to GPT to suggest a new opinion score for an agent.

Role prompt: Pretend you are a `«age»`-year old `«race»` `«sex»` who has completed `«education»`
Main prompt: Pretend you have a political opinion score `«opinion»` where -10 is far-right Republican and 10 is far-left Democrat. What is your new opinion score after you see `«message»` sent to you on social media? Do not explain your reasoning.

For replicability, note that a simulation primarily depends on OpenAI 1.58.1 (for the GPT API), Numpy 1.26.4 to store the agents' attributes and connections (scaling-up since computations with arrays are faster than Python primitive data structures), and re 2.2.1 (to parse the LLM response using regular expression).

3.2 Empirical data

Social network. The far right heavily relies on social media, with an established presence on all well-known platforms including Twitter [29]. The use of such a mainstream platform is suitable for our case study, since we simulate the potential spread of far-right ideas among a *general population* rather than only among right-leaning groups who may already endorse some of these ideas (e.g., on the Parler or Truth Social platforms). To support replicability of our simulations, we use a public domain sample of Twitter data consisting of $11,316,811$ nodes (users) and $85,331,846$ edges (representing a follower relationship) [2].

While using the *whole sample* for simulations is sometimes unnecessary to observe representative trends, it is potentially very expensive [20] and wasteful

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_16

in terms of computations, and costly when using paid LLMs such as GPT. Simulations commonly right-size the computations by using a *sufficient sample* size for the experiments [34] (see subsection 4.2). Each sample should be representative of the data. Using a python library for graph sampling [49], we noted that a random sampling of node was unsuitable (the network consists of scattered users lacking connectivity for message spread) and sampling by PageRank had poor scaling (memory needs exceed 32GB). We use a *degree-based sampling* strategy [26], which preserves characteristics of the degree distribution (e.g., heavily skewed) but creates changes as evidenced by a Jensen-Shannon distance of 0.64 between a sample of 200 nodes and the whole network (using normalized degree distributions). As noted by Moran-Tovar and colleagues, "while the degree distribution ignores the specific topology of the network, it captures the effect of largely connected nodes or hubs on the transmission statistics" [38]. It is thus appropriate to study the dynamics of spreading phenomena, but it may not support a more fine-grained analysis (e.g., detecting communities). The sampling library returns an *undirected* graph but Twitter data is *directed*: a user A follows B and that is not necessarily reciprocal. We restored directionality by checking for all sample nodes whether they appeared as a pair the edge list of the original data, then we added the corresponding edges.

Messages. While Twitter shows us the *reaction* of a wide population to extreme ideas, the diversity of topics and tones encountered on this platform limits its usability as a *source* for such ideas. The partisan tone of alternative media websites make them valuable sources to retrieve far-right ideas. In particular, Truth Social (launched by Donald Trump) is the most right-leaning alternative platform [17]. Per the Pew Research Center, most prominent accounts (94%) were individuals rather than organizations [17]. It is thus a suitable source to collect far-right messages as expressed by individuals. For replicability, we use an open Truth Social dataset crawled from February 21st, 2022 to October 15th, 2022. The authors sampled posts and accounts using Trump's account as the seed, then spread to his followers and other popular accounts. The dataset has over $823,000$ posts and over $454,000$ accounts [22].

Not all messages can be used to spread ideas via a simulation, since some may be too short or are only interpretable in the context of a discussion. We thus undertook three typical steps for pre-processing political online messages [50]. First, we used the well established `TextBlob` Python library 0.18.0 [3] to assign a sentiment polarity to each message from -1 (very negative) to 1 (very positive). In order to study reactions regarding strongly worded messages, we retained posts with absolute polarity above 0.9. Then, we removed posts with fewer than 25 characters (since they are either too context-dependent or lack content). Finally, we removed all posts that contained links because the LLM may attempt and fail to retrieve the URL's content thus causing simulation errors. This technical limitation may impact the results since real-world social media users may be more affected by posts that include references. Following our three steps, we obtained $13,239$ messages to spread in our simulations (exemplified in Box 2).

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_16

Box 2. Sample of Truth Social posts

- Insurrection my ass. Video after video now showing the Capitol Police encouraging and inviting people through! We have all seen the videos and most of us have them saved!!!
- Ilhan Omar is the perfect example of why we need an immigration moratorium in America.
- Ignorant idiots! Y'all have no idea what the CDC, WHO, and this government is getting you to do! Wake up now!
- If the U.S can afford to send 40 Billion dollars to Ukraine, then We can afford to put armed security in all 131,000 schools in America!! Protect the Children!! Evil people do not care about laws!!
- Yup! Fauci is evil and does need to be locked up!

Post frequency. Social media accounts do not continuously produce content; rather, they post at a given frequency. To seed their simulated Twitter network, Ben Sliman and Kohli analyzed the empirical distribution of the average number of tweets per day as a function of the number of followers [7]. We extracted the numerical data from the plotted distribution in their article using https://automeris.io Since the distribution is discrete (e.g., 260 followers, 270, 280...), we used a linear interpolation to obtain a continuous distribution that allows us to quantify the frequency for all user accounts (e.g., 263 followers). The authors' plotted distribution starts at 24 followers, thus we used an extrapolation to estimate the posting frequency of users with fewer followers, under the assumption that agents without followers would have no posts.

Demographics. Reviews on conspiracies and politically divisive decisions in the U.S. (e.g., whether to vaccinate for COVID-19) have shown that key determinants include age, sex, educational attainment, race and ethnicity. As we previously detailed, *these four determinants should be initialized together when creating virtual agents due to their dependencies* [6]. Otherwise, we would erroneously create agents with plausible age distribution and (separately) valid educational attainments, but their joined distributions may not match the data. We use the U.S. Census CPS Basic Monthly Data from October 2024 as this nationally representative survey provides tables for the joint distributions of the four social determinants [55].

4 Results

Evaluations of LLM agents can be performed at two levels [20]. At the *micro-level*, simulated decisions from the agents (particularly through the prism of the LLM) must align with real-world data (subsection 4.1). At the *macro-level* we assess dynamics over the entire population, which may differ from the sum of the individuals given that ABMs often show *emerging* properties. Prior LLM/ABM

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_16

models for the spread of political opinions used 15 agents over 10 simulation steps and presented macro-level findings through visuals [62]. However, there is a risk for such results to be an artifact of the limited model's size, or that findings lack statistical significance. We thus use a statistical approach to identify a suitable model size (subsection 4.2) then we analyze the results (subsection 4.3).

4.1 Validating the use of LLMs for political opinions

We use a LLM to quantify the political opinion of an agent based on four demographic features. Assessing the accuracy of the LLM in performing this complex task contributes to validating its use to initialize our agents and informs us of the confidence margin associated with the simulation results. Since the demographics of voters change over time, we compared the LLM results with the most recent 2022 data from the Pew Research Center data [25] as well as the 2024 post-voting polls from NBC News [41]. Surveys have limited generalizibility since many eligible voters do not vote. *On average, the prediction of GPT are 6 percentage points away from either of the two surveys*, which makes it suitable for our application. The predictions are more accurate with respect to race (maximum error of 6%), followed by age (overestimating elderly as conservative and young adults as liberal), educational attainments (overestimating liberals among college graduates) and sex (with the highest error of 16% on female voters). Our results based on GPT-4 confirm previous reports based on GPT-3 [4]: it could not be distinguished from humans when associating keywords with political parties; it was also highly correlated with political votes when agents profiles were provided based on race, age, sex, and seven other characteristics.

Table 1. Prevalence of right-leaning voters in two surveys vs. prediction of GPT. For GPT, we create a complete population and we aggregate to obtain the target feature value. For example, for 'male', we aggregate all male agents with weights corresponding to the prevalence of race, educational attainments, and age category among males. We also generate American Indians and Alaskan Natives (as one group) but they were omitted from the racial breakdown due to their low prevalence.

Demographic feature	Group	Pew 2022	NBC 2024	GPT-4
Sex	Male	54%	55%	65%
	Female	48%	45%	32%
Race	White	57%	57%	56%
	Black	5%	13%	10%
	Asian	32%	40%	34%
Educational attainments	Postgraduate	37%	38%	45%
	College graduate	48%	50%	36%
	Some college	54%	51%	49%
	High school or less	59%	62%	60%
Age category	18-29	31%	43%	32%
	30-49	45%	47%	43%
	50-64	55%	54%	54%
	65+	56%	50%	62%

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_16

4.2 Right-sizing the model: effects of scaling

We aim to identify a *sufficient* population size so that results reflect the dynamics of the model instead of being an artifact of the model's size (see section 3.2–**Social network**). As in previous works, we identify a sufficient size by starting with a small population, gradually increasing its size, and measuring whether the outputs depend on the population size (cf. Figures 7–9 in [24]). As expected, the standard deviation decreases as the population size increases (Table 2). A one-way ANOVA between the simulation outputs for each population size shows no statistically significant differences for 11 of the 13 demographic groups. In the case of black agents (ANOVA p-value=0.03), a post hoc Tukey HSD revealed that a population size of 100 was statistically different from 150 (Q=4.57, p=0.02); there were no other differences. In the case of college graduates (ANOVA p-value=0.01), a population size of 100 was statistically different from 150 agents (Q=5.23, p=0.009) and 50 agents (Q=4.12, p=.04); again, there were no differences between other sizes. In summary, a population size of 100 is insufficient as its results differ from other sizes. There is no difference between population sizes of 150 and 200 so either can be employed. In the remainder of this paper, we use 200 agents as it yields a narrower standard deviation.

4.3 Dynamics of the population

Prior works measured the diversity of opinions as the *number* of different opinions among the agents [62], but this does not account for the *frequency* at which these opinions hold. We thus measure the diversity of opinions using Shannon

Table 2. For each population size, we report the average ± standard deviation over 5 runs of the political opinion score, which ranges from -10 (far-right) to 10 (far-left). We also perform an ANOVA on outputs of the 5 runs across population sizes.

		Simulated population size (number of agents)				ANOVA
Demog.	**Group**	***50***	***100***	***150***	***200***	p-value
Sex	Male	-4.04 ± 1.09	-3.20 ± 1.03	-3.72 ± 0.44	3.42 ± 0.32	0.39
	Female	-1.24 ± 0.75	-0.20 ± 1.26	-1.48 ± 0.39	1.17 ± 0.52	0.09
Race	White	-3.33 ± 0.88	-2.31 ± 0.92	-3.37 ± 0.48	3.04 ± 0.25	0.09
	Black	1.40 ± 0.70	1.97 ± 0.60	0.63 ± 0.77	-1.14 ± 0.54	0.03
	Asian	-2.18 ± 2.17	-1.08 ± 0.42	-1.62 ± 1.79	0.38 ± 1.30	0.33
Educational attainments	Postgrad.	-0.56 ± 1.62	-0.56 ± 1.46	-0.60 ± 0.91	0.55 ± 0.52	0.99
	College grad.	-1.69 ± 0.90	-0.53 ± 0.66	-2.00 ± 0.46	1.36 ± 0.32	0.01
	Some college	-2.28 ± 1.85	-0.83 ± 1.56	-2.67 ± 0.64	2.36 ± 0.85	0.16
	≤high school	-4.29 ± 0.98	-3.65 ± 0.90	-3.86 ± 0.56	3.66 ± 0.25	0.48
Age cat.	18-29	-0.90 ± 0.83	0.07 ± 1.02	-0.56 ± 0.26	0.28 ± 0.45	0.21
	30-49	-1.34 ± 1.45	-1.50 ± 1.15	-2.30 ± 0.72	1.93 ± 0.17	0.44
	50-64	-3.21 ± 0.78	-2.56 ± 0.65	-3.16 ± 0.57	2.95 ± 0.84	0.47
	65+	-4.07 ± 2.07	-2.66 ± 1.15	-4.24 ± 0.88	3.96 ± 0.43	0.23
Standard deviation		1.51	1.23	0.86	0.60	

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_16

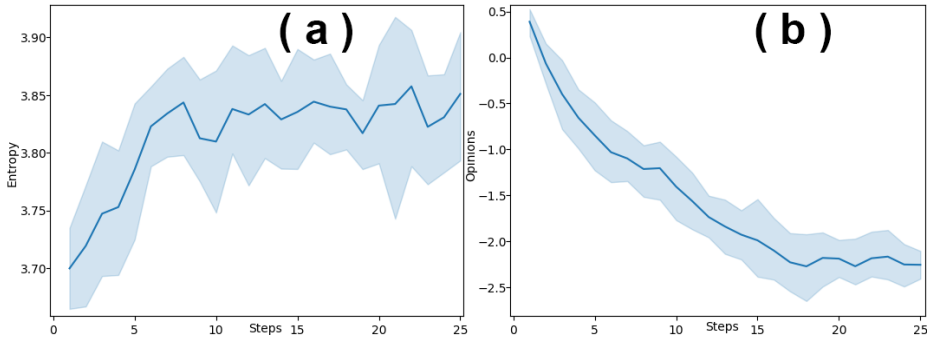

Fig. 1. The diversity of opinions quickly plateaus (a) while the average opinion plateaus after 20 steps (b). Standard deviations (blue bands) are based on 5 simulation runs.

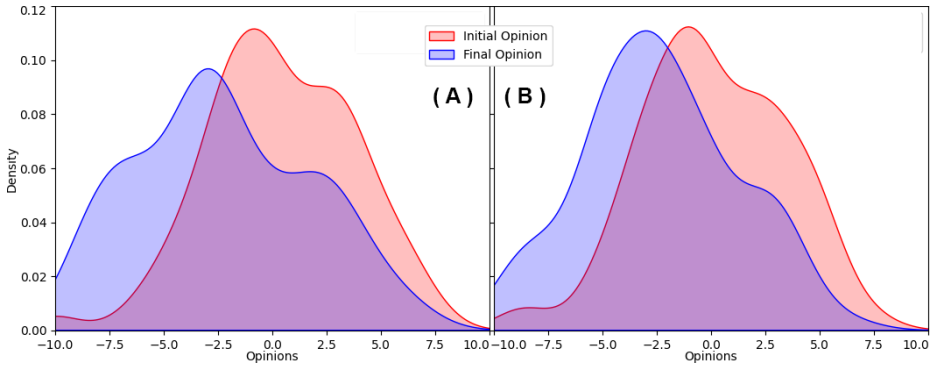

Fig. 2. The initial distribution of opinions at $t = 0$ morphs into its final configuration at $t = 25$ across two simulation runs (A, B) for 200 agents.

entropy, where a higher entropy means more diversity. Figure 1-a shows that the diversity initially rises modestly from 3.70 and plateaus at 3.84. The average opinion value starts almost neutral in the population and steadily becomes more right leaning, oscillating at -2.25 (Figure 1-b). Together, these results suggest that being *exclusively* exposed to far-right messages produces a change in the population. As shown in Figure 2, this change is not merely a *shift* where every individuals experience the same decrease in opinion score. Rather, individuals react based on their demographics, with some holding firmly to their opinions (as the distribution continues to go up to 10 – far left) and others having a stronger reaction (as shown by the increased weight in the first half of the distribution).

5 Discussion

Given prior works on using GPT to emulate voting patterns [4] or key political debates such as abortion [62], we have shown the feasibility of simulating *changes*

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_16

in opinions due to exposure to specific political messages. Our work confirms the potential of combining LLMs with ABMs to to develop models that represent human behavior and decision-making [23]. As stated by Park and colleagues, a model *generates* behaviors in social media according to certain specifications [44]. The goal of a model is not always to merely witness a phenomenon (that we already knew was happening); rather, it serves as a virtual laboratory to test the consequences of possible interventions. The reddit simulation from Park *et al.* thus paired a generating model with a what-if component to study scenarios such as moderator interventions. By following our process or directly reusing our open-source implementation, researchers can test strategies such as combining the model with detection algorithms (e.g., for hate speech, incitement to violence, or misinformation) to delete posts, ban their authors, or algorithmically deprioritize posts (i.e., reduce their visibility in a reader's feed).

As we are only in the infancy of generative agent-based modeling, there are several interesting avenues to extend the model. Pastor-Galindo *et al.* stressed that it is "imperative to simulate and model realistic social networks" by modeling three aspects [45]. In this paper, we focus on the first aspect of *direct communications* as agents write posts to which their followers react. Our model did not account for *information sharing* by mining links and other dynamic contents shared on a network (which can be achieved by the LLM), and we did not represent how *user engagement* varies depending on the type of content (which needs a change in the model and prompt). Dynamicity can be important depending on the simulated time window: at the scale of a few days, we can assume that the network is static (as in our study), but over longer durations, there would be changes since individuals unfollow others or create links.

While our study used several real-world datasets for empirical grounding, there are two limitations in data availability regarding individual opinions in general (see section 2.1) and for certain political groups in particular. First, our agents have a representative and internally consistent set of demographics but given the paucity of data that *associates such features with social media accounts*, agents were assigned at random to Twitter accounts. This makes it possible for an extremely left-leaning agent to follow extreme-right accounts. As a result, the changes observed in our simulation are an over-estimate of changes happening in real-world networks characterized by assortativity and echo chambers. Second, we examined changes due to exposure to far-right posts from Truth Social, while noting that social media platforms such as Twitter/X contain a *variety* of posts. At present, there is no left-wing equivalent to Truth Social that allows for the same large-scale data collection. As new platforms (e.g., Bluesky) emerge, it may become possible to simulate complete exposure to left- and right-leaning posts.

Acknowledgments. We gratefully acknowledge the financial support of American University's Signature Research Initiative Project. SZ wishes to thank Eric Schuler for support in using American University's High Performance Computing Cluster Zorro.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_16

References

1. Abelson, R.P.: Mathematical models of the distribution of attitudes under controversy. Contributions to mathematical psychology (1964)
2. Aché, M.: Twitter edge nodes, https://www.kaggle.com/datasets/mathurinache/twitter-edge-nodes/data, last accessed 12/31/24
3. Aljedaani, W., Rustam, F., Mkaouer, M.W., et al.: Sentiment analysis on twitter data integrating textblob and deep learning models: The case of us airline industry. Knowledge-Based Systems **255**, 109780 (2022)
4. Argyle, L.P., Busby, E.C., Fulda, N., et al.: Out of one, many: Using language models to simulate human samples. Political Analysis **31**(3), 337–351 (2023)
5. Baele, S.J., et al.: Uncovering the far-right online ecosystem: An analytical framework and research agenda. Stud. Confl. Terror. **46**(9), 1599–1623 (2023)
6. Beerman, J.T., Beaumont, G.G., Giabbanelli, P.J.: A scoping review of three dimensions for long-term covid-19 vaccination models: hybrid immunity, individual drivers of vaccinal choice, and human errors. Vaccines **10**(10), 1716 (2022)
7. Ben Sliman, M., Kohli, R.: Asymmetric relations and the friendship paradox. Columbia Business School Research Paper (18-73) (2018)
8. von Briesen, E.M., Bacaksizlar, N.G., Hadzikadic, M.: Modeling genocide at the system and agent levels. Journal on Policy and Complex Systems **3**(2), 31–48 (2017)
9. Brown, O., Smith, L.G., Davidson, B.I., et al.: Online signals of extremist mobilization. Personality and Social Psychology Bulletin p. 01461672241266866 (2024)
10. Cointet, J.P., Roth, C.: How realistic should knowledge diffusion models be? JASSS **10**(3), 5 (2007)
11. Coscia, M., Rossi, L.: How minimizing conflicts could lead to polarization on social media: An agent-based model investigation. PloS one **17**(1), e0263184 (2022)
12. Crokidakis, N.: Recent violent political extremist events in brazil and epidemic modeling: The role of a sis-like model on the understanding of spreading and control of radicalism. International Journal of Modern Physics C **35**(02), 2450015 (2024)
13. Deffuant, G., Neau, D., Amblard, F., Weisbuch, G.: Mixing beliefs among interacting agents. Advances in Complex Systems **3**(01n04), 87–98 (2000)
14. DeGroot, M.H.: Reaching a consensus. Journal of the American Statistical association **69**(345), 118–121 (1974)
15. Ebner, J., Kavanagh, C., Whitehouse, H.: Assessing violence risk among far-right extremists: A new role for natural language processing. Terrorism and political violence **36**(7), 944–961 (2024)
16. Flache, A., Mäs, M., Feliciani, T., et al.: Models of social influence: Towards the next frontiers. JASSS **20**(4), 2 (2017)
17. Forman-Katz, N., Stocking, G.: Key facts about truth social (2022)
18. Froio, C., Ganesh, B.: The transnationalisation of far right discourse on twitter: Issues and actors that cross borders in western european democracies. European societies **21**(4), 513–539 (2019)
19. Gambaro, J.P., Crokidakis, N.: The influence of contrarians in the dynamics of opinion formation. Physica A **486**, 465–472 (2017)
20. 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)
21. Geller, M., Vasconcelos, V.V., Pinheiro, F.L.: Toxicity in evolving twitter topics. In: International Conference on Computational Science. pp. 40–54. Springer (2023)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_16

22. Gerard, P., Botzer, N., Weninger, T.: Truth social dataset. In: Proc. Int. AAAI Conf. on Web and Social Media. vol. 17, pp. 1034–1040 (2023)
23. Ghaffarzadegan, N., Majumdar, A., et al.: Generative agent-based modeling: an introduction and tutorial. System Dynamics Review **40**(1), e1761 (2024)
24. Gibson, M., Portugal Pereira, J., et al.: Agent-based modelling of future dairy and plant-based milk consumption for uk climate targets. JASSS **25**(2) (2022)
25. Hannah Hartig, Andrew Daniller, S.K., Green, T.V.: (2023), pewresearch.org/politics/2023/07/12/voting-patterns-in-the-2022-elections
26. Hu, P., Lau, W.C.: A survey and taxonomy of graph sampling. arXiv preprint arXiv:1308.5865 (2013)
27. Huszár, F., Ktena, S.I., O'Brien, C., Belli, L., Schlaikjer, A., Hardt, M.: Algorithmic amplification of politics on twitter. PNAS **119**(1), e2025334119 (2022)
28. Jager, W., Amblard, F.: Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change. Comput. Math. Organ. Theory **10**, 295–303 (2005)
29. Kakavand, A.E.: Far-right social media communication in the light of technology affordances: a systematic literature review. Annals of the International Communication Association **48**(1), 37–56 (2024)
30. Kasaie, P., Kelton, W.D.: Guidelines for design and analysis in agent-based simulation studies. In: Winter Simulation Conference. pp. 183–193 (2015)
31. Laver, M., Sergenti, E.: Party Competition: An Agent-Based Model. Princeton University Press (2012)
32. Lipiecki, A.: Strategic promotional campaigns for sustainable behaviors: Maximizing influence in competitive complex contagions. In: International Conference on Computational Science. pp. 62–70. Springer (2024)
33. Lorenz, J.: Continuous opinion dynamics under bounded confidence: A survey. International Journal of Modern Physics C **18**(12), 1819–1838 (2007)
34. Lutz, C.B., Giabbanelli, P.J.: When do we need massive computations to perform detailed covid-19 simulations? Adv. Theory Simul. **5**(2), 2100343 (2022)
35. Mahmoudi, A., Jemielniak, D., Ciechanowski, L.: Echo chambers in online social networks: a systematic literature review. IEEE Access (2024)
36. Mead, E.L., McNerney, H.W., Agarwal, N.: Text mining domestic extremism topics on multiple social media platforms. In: Int. Conf Computational Linguistics and Natural Language Processing. pp. 104–111. IEEE (2024)
37. Memmert, L., Cvetkovic, I., Bittner, E.: The more is not the merrier: Effects of prompt engineering on the quality of ideas generated by gpt-3. In: Proc. 57th Hawaii International Conference on System Sciences. pp. 7520–7529 (2024)
38. Morán-Tovar, R., Gruell, H., et al.: Stochasticity of infectious outbreaks and consequences for optimal interventions. J. Physics A **55**(38), 384008 (2022)
39. Müller, A., Lopez-Sanchez, M.: Countering negative effects of hate speech in a multi-agent society. In: Artificial Intelligence R&D, pp. 103–112 (2021)
40. Mustafee, N., Powell, J.H.: From hybrid simulation to hybrid systems modelling. In: 2018 Winter Simulation Conference (WSC). pp. 1430–1439. IEEE (2018)
41. NBC News: Exit polls (2024), https://www.nbcnews.com/politics/2024-elections/exit-polls
42. Negahban, A., Giabbanelli, P.J.: Hybrid agent-based simulation of adoption behavior and social interactions: Alternatives, opportunities, and pitfalls. IEEE Transactions on Computational Social Systems **9**(3), 770–780 (2021)
43. O'Hara, K., Stevens, D.: Echo chambers and online radicalism: Assessing the internet's complicity in violent extremism. Policy & Internet **7**(4), 401–422 (2015)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_16

44. Park, J.S., Popowski, L., Cai, C., et al.: Social simulacra: Creating populated prototypes for social computing systems. In: Proc. 35th Annual ACM Symposium on User Interface Software and Technology. pp. 1–18 (2022)
45. Pastor-Galindo, J., Nespoli, P., Ruipérez-Valiente, J.A.: Large-language-model-powered agent-based framework for misinformation and disinformation research: Opportunities and open challenges. IEEE Security & Privacy (2024)
46. Popa-Wyatt, M.: Online hate: Is hate an infectious disease? is social media a promoter? Journal of Applied Philosophy **40**(5), 788–812 (2023)
47. Rainer, H., Krause, U.: Opinion dynamics and bounded confidence: models, analysis and simulation. JASSS **5**(3) (2002)
48. Rothut, S., Schulze, H., et al.: Ambassadors of ideology: A conceptualization and computational investigation of far-right influencers, their networking structures, and communication practices. New Media & Society **26**(12), 7120–7147 (2024)
49. Rozemberczki, B., Kiss, O., Sarkar, R.: Little ball of fur: a python library for graph sampling. In: Proc. 29th ACM Int. Conf. information & knowledge management. pp. 3133–3140 (2020)
50. Sandhu, M., Vinson, C.D., Mago, V.K., Giabbanelli, P.J.: From associations to sarcasm: mining the shift of opinions regarding the supreme court on twitter. Online Social Networks and Media **14**, 100054 (2019)
51. Serrano, E., Iglesias, C.Á., Garijo, M.: A novel agent-based rumor spreading model in twitter. In: Proc. 24th Int. Conf. on World Wide Web. pp. 811–814 (2015)
52. Stokes, B.M., Jackson, S.E., Garnett, P., Luo, J.: Extremism, segregation and oscillatory states emerge through collective opinion dynamics in a novel agent-based model. The Journal of Mathematical Sociology **48**(1), 42–80 (2024)
53. Tao, K., et al.: Gpt-4 performance on querying scientific publications: reproducibility, accuracy, and impact of an instruction sheet. BMC Medical Research Methodology **24**(1), 139 (2024)
54. Ting, M.M., Bendor, J., Diermeier, D., Siegel, D.A.: A behavioral theory of elections. Princeton University Press (2011)
55. United States Census Bureau: CPS Basic Monthly (2024 OCT) Custom Table on PESEX, PTDTRACE, PEEDUCA, PRTAGE (2024), https://data.census.gov/app/mdat/CPSBASIC202410
56. Unlu, A., Kotonen, T.: Online polarization and identity politics: An analysis of facebook discourse on muslim and lgbtq+ communities in finland. Scandinavian Political Studies **47**(2), 199–231 (2024)
57. Waldherr, A., et al.: Worlds of agents: Prospects of agent-based modeling for communication research. Commun. Methods Meas. **15**(4), 243–254 (2021)
58. Weisburd, D., Wolfowicz, M., et al.: Using agent based modelling to advance evaluation research in radicalization and recruitment to terrorism: Prospects and problems. Stud. Confl. Terror. pp. 1–24 (2024)
59. Westermann, C.J., Coscia, M.: A potential mechanism for low tolerance feedback loops in social media flagging systems. Plos one **17**(5), e0268270 (2022)
60. Zhang, H., Vorobeychik, Y.: Empirically grounded agent-based models of innovation diffusion: a critical review. Artificial Intelligence Review **52**, 707–741 (2019)
61. Zhang, X., Davis, M.: E-extremism: A conceptual framework for studying the online far right. New Media & Society **26**(5), 2954–2970 (2024)
62. Zheng, W., Tang, X.: Simulating social network with llm agents: An analysis of information propagation and echo chambers. In: International Symposium on Knowledge and Systems Sciences. pp. 63–77. Springer (2024)
63. Zimmerman, F., Bailey, D.D., Muric, G., et al.: Attraction to politically extreme users on social media. PNAS nexus **3**(10), pgae395 (2024)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_16