Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

AURORA: ENHANCING SYNTHETIC POPULATION REALISM THROUGH RAG AND SALIENCE-AWARE OPINION MODELING

Rebecca Marigliano[1] and Kathleen M. Carley[1]

[1]Societal and Software Systems Department, Carnegie Mellon University, Pittsburgh, USA

ABSTRACT

Simulating realistic populations for strategic influence and social-cyber modeling requires agents that are demographically grounded, emotionally expressive, and contextually coherent. Existing agent-based models often fail to capture the psychological and ideological diversity found in real-world populations. This paper introduces AURORA, a Retrieval-Augmented Generation (RAG)-enhanced framework that leverages large language models (LLMs), semantic vector search, and salience-aware topic modeling to construct synthetic communities and personas. We compare two opinion modeling strategies and evaluate three LLMs—gemini-2.0-flash, deepseek-chat, and gpt-4o-mini—in generating emotionally and ideologically varied agents. Results show that community-guided strategies improve meso-level opinion realism, and LLM selection significantly affects persona traits and emotions. These findings demonstrate that principled LLM integration and salience-aware modeling can enhance the realism and strategic utility of synthetic populations for simulating narrative diffusion, belief change, and social response in complex information environments.

1 INTRODUCTION

Understanding how opinions and emotions spread online requires tools that can realistically simulate belief formation, sentiment expression, and social identity dynamics. Traditional agent-based models (ABMs) often rely on simplistic rules, failing to capture the cognitive, cultural, and emotional complexity of real populations. Advances in large language models (LLMs) and retrieval-augmented generation (RAG) offer a path forward by enabling the creation of synthetic agents that are demographically grounded, contextually informed, and psychologically coherent.

We introduce AURORA (AI-Utilized Retrieval for Optimized Representation of Audiences), a RAG-enhanced ABM framework for generating diverse, realistic synthetic populations. AURORA combines semantic vector search, salience-aware topic modeling, and LLM-driven persona construction to simulate belief dynamics across multiple scales—from national discourse to individual emotional traits.

To evaluate AURORA, we model Taiwan's information environment across three salient topics, comparing two opinion modeling strategies—Community-Guided Opinion Assignment (CGO) and Persona-Differentiated Strategy (PDS)—and three LLMs. We define realism as the extent to which synthetic agents exhibit demographic plausibility, emotional coherence, and ideological diversity. We assess this through opinion variation across social groups, emotion-personality alignment, and salience-based opinion spread, capturing how strongly held beliefs resist change in high-salience contexts.

Research Questions:

1. **RQ1:** How do different opinion modeling strategies (CGO & PDS) affect the realism and variability of simulated opinions across national, community, and persona levels?
2. **RQ2:** To what extent do LLMs vary in their generation of psychological traits and emotional states, and how does this variation influence the construction of synthetic agents?

3. **RQ3:** Can salience-aware, RAG-grounded generation pipelines improve the alignment of synthetic population behavior within discourse structures and ideological patterns?

Our results demonstrate that salience-aware, RAG-enhanced generation improves the contextual realism and ideological coherence of synthetic populations. Community-guided opinion modeling better captures sub-national diversity, while the choice of LLM significantly shapes psychological and emotional traits in personas. These findings highlight the importance of aligning opinion modeling strategy and LLM selection with simulation goals—whether to model polarization, disengagement, or emotionally expressive behaviors in social-cyber scenarios.

2 RELATED WORKS

2.1 Simulating Human Behavior with Language Models

Recent work explores the use of LLMs to simulate social behavior. Park et al. model emergent interactions through generative agents with memory Park et al. 2023, while Avery et al. use persona-conditioned prompting to simulate survey responses Argyle et al. 2023. Liu et al. 2023 and Si et al. 2024 simulate audience feedback and opinion shifts using structured prompting. However, these approaches often rely on static, homogeneous agent behavior with limited contextual grounding. Most lack mechanisms for temporal updates, salience modulation, or integration of real-world data. As Ribeiro notes, simulations built soley on LLM output risk epistemic uniformity and reduced realism Ribeiro 2025.

2.2 Retrieval-Augmented Generation (RAG)

Retrieval-Augmented Generation (RAG) blends LLMs with external knowledge retrieval to improve factual grounding and adaptability, as first demonstrated by Lewis et al using dense document retrieval for open-domain QA Lewis 2020. While toolkits like LangChain extend RAG's utility through modular pipelines LangChain Team 2023, most implementations remain tailored to static task completion rather than dynamic simulation. Challenges persist in applying RAG to behavioral modeling, including alignment of retrieved content with psychologically coherent agent states, sensitivity to corpus design and embedding strategies, and a general lack of salience-aware or temporally adaptive retrieval. Existing systems often treat retrieved context as fixed input, limiting their capacity to simulate evolving beliefs and attention in agents.

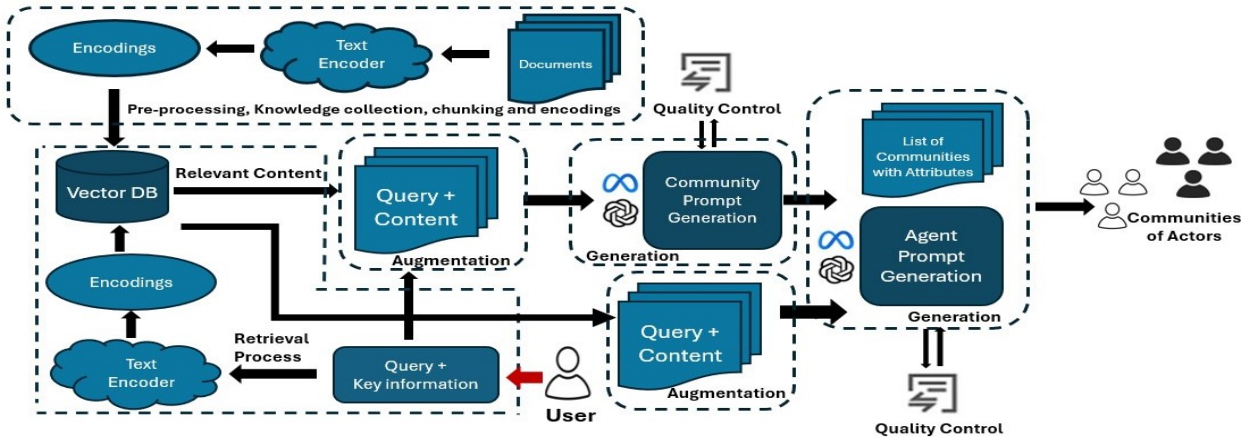


Figure 1: The AURORA model workflow from semantic encoding and vector database creation through context retrieval, community and agent prompt generation, to the final synthesis of realistic synthetic communities and personas.

3 AURORA MODEL ARCHITECTURE

The AURORA framework is structured as an advanced Retrieval-Augmented Generation (RAG)-enhanced Agent-Based Model (ABM). At its core, AURORA integrates semantic vector representations, retrieval methods, and Large Language Models (LLMs) to generate detailed, realistic synthetic personas within defined community contexts. The process involves distinct stages: semantic data embedding, community creation, persona generation, and dynamic simulation.

3.1 Semantic Vector Database and RAG

AURORA uses a high-dimensional semantic vector database, implemented via ChromaDB, to support persona and community generation through Retrieval-Augmented Generation (RAG). A curated corpus of country-specific documents, social discourse, and demographic data is preprocessed by chunking text, removing duplicates, and adding metadata. Each chunk c_i is embedded into a vector $\mathbf{v}_i \in \mathbb{R}^d$ using a configurable encoder f_{encoder}, such as `OpenAIEmbeddings`, forming a vector store $V = \{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$. For a query q, AURORA computes the embedding $\mathbf{q} = f_{\text{encoder}}(q)$, and retrieves the top-$k$ most similar vectors using cosine similarity:

$$\text{sim}(\mathbf{q}, \mathbf{v}_i) = \frac{\mathbf{q} \cdot \mathbf{v}_i}{\|\mathbf{q}\| \|\mathbf{v}_i\|}.$$

3.2 Integration with Community and Persona Generation

The retrieved content from the semantic vector database forms the contextual backbone for both community construction and persona generation. After relevant document chunks are selected and formatted via the RAG process, this context is embedded directly into standardized LLM prompt templates through the `relevant_text` field. This ensures that generated outputs are not only syntactically coherent but also semantically grounded in real-world discourse.

3.2.1 Community Formation and Salience Assignment

For each community C_i, AURORA formulates a semantic query based on community name and country of origin (e.g., `"community X in country Y"`). This query is used to retrieve the top-k relevant textual segments from the vector database. It is inserted into a structured prompt template that guides the LLM to generate a detailed community profile.

Using the retrieved content and configured scenario parameters (e.g., number of communities, community categories, key topics), the LLM generates standardized outputs that define:

- Community name and purpose
- Demographic structure (age, gender distribution, etc.)
- Political leaning (continuous and labeled)
- Topic list $\{t_1, t_2, ..., t_k\}$ with associated salience scores

Each community is also associated with a category (e.g., labor union, student group, activist coalition), which influences its structural composition and ideological baseline. The output is validated using standardized parsers and stored as a 'Community' object within the simulation environment.

Topic Salience Modeling. For each key topic t_k associated with community C_i, the LLM first generates a salience label (e.g., Low, Medium, High, Very High) based on both the scenario context and retrieved text. This label is then mapped to a corresponding salience score $S_k \in [0, 1]$, reflecting the relative importance of the topic to that community.

Salience values influence downstream persona modeling, including: variance in opinion stance sampling, likelihood of topic engagement, and emotional intensity during expression or interaction.

Community Assembly. The full set of synthetic communities $\mathscr{C} = \{C_1, C_2, ..., C_n\}$ is assembled by repeatedly invoking the RAG+LLM pipeline, guided by a configuration-defined number of communities and categories. Each community is generated independently, allowing for modular diversity and scalability.

This formation process ensures that all communities are contextually grounded, demographically plausible, and ideologically distinct—forming the foundational layer for subsequent persona generation and social simulation.

3.2.2 Persona Generation with Emotional and Opinion States

Once communities have been generated, AURORA constructs individualized synthetic personas that inherit demographic and ideological characteristics from their associated community, while introducing psychological and behavioral diversity at the individual level. Persona generation is carried out through an ordered, RAG-informed LLM pipeline that ensures consistency with upstream community data while enabling population-scale heterogeneity.

Persona Generation Process. For each community C_i, AURORA synthesizes a set of personas using a standardized language model pipeline that integrates community profiles with contextual information. This approach produces persona profiles that capture the essential demographic and ideological traits of the community while ensuring overall coherence. Each persona is uniquely identified and stored with relevant metadata.

4 COMMUNITY-LEVEL OPINION MODELING AND SALIENCE-DRIVEN TOPIC ALIGNMENT

AURORA models ideological stance distributions by hierarchically aligning country-level, community-level, and persona-level opinions. At the core of this system is a salience-aware, multi-resolution opinion generation pipeline that captures micro-level psychological variability.

Each synthetic community C_i is assigned a set of opinions over topics $t_k \in \mathscr{T}$, where $\mathscr{T}$ is the global topic set for a given simulation scenario. Topic-specific opinions O_k are generated using topic- and community-specific queries. These contexts ground structured LLM prompts that output both Cabrera et al. 2021:

- An **opinion value** $O_k \in [-2, 2]$, where polarity and strength of sentiment are encoded,
- A **salience value** $S_k \in [0, 1]$, representing the issue's centrality to the community Dong et al. 2018.

Formally, the community-level opinion on topic t_k is represented as:

$$O_k, S_k = f_{\text{LLM}}(t_k, C_i, \mathscr{D}_{\text{context}}), \tag{1}$$

where f_{LLM} is the structured language model chain, and $\mathscr{D}_{\text{context}}$ is retrieved contextual information.

4.1 Salience-Weighted Distributional Modeling of Persona Opinions

Once the community-level opinions O_k and salience values S_k are established, AURORA samples persona-level opinions by introducing stochastic variance modulated by topic salience. For persona P_j in community C_i, the opinion on topic t_k is given by Alizadeh and Cioffi-Revilla 2016, Cabrera et al. 2021:

$$O_{p,j}^{(k)} \sim \begin{cases} \mathscr{N}(O_k, \sigma) & \text{if } P(O_k) \text{ is normal,} \\ \text{Skewed}(O_k) & \text{if } P(O_k) \text{ is skewed,} \\ \text{Bimodal}(O_k) & \text{if } P(O_k) \text{ is polarized,} \end{cases} \tag{2}$$

A normal distribution is used when the topic exhibits a neutral or balanced national stance, reflecting general consensus. When the population shows an asymmetric stance—such as when one side dominates the discourse but a minority voice persists—a skewed distribution is employed, implemented via parameterized

Beta distributions Alizadeh and Cioffi-Revilla 2016. For topics characterized by strong ideological duality and polarization, a bimodal distribution is used, capturing the presence of two dominant yet opposing viewpoints Alizadeh and Cioffi-Revilla 2016.

The standard deviation σ of each distribution is inversely proportional to topic salience:

$$\sigma = \sigma_0 \cdot (1 - S_k), \tag{3}$$

where σ_0 is a predefined upper bound (typically 0.5 to 1.5). High-salience topics (e.g., identity politics, religion) lead to tighter clustering around the community stance, while low-salience topics (e.g., niche policies) produce higher variance and weaker alignment.

4.2 From Community to Persona: Salience-Guided Opinion Synthesis

Each persona P_j's opinion on topic t_k is generated by combining the community opinion O_k, topic salience S_k, and psychologically plausible noise ε_j, sampled from a salience-aware distribution:

$$O_{p,j}^{(k)} = \text{Clip}(O_k + \varepsilon_j), \quad \varepsilon_j \sim \mathscr{D}(O_k, S_k)$$

where the standard deviation of ε_j is inversely proportional to the salience S_k, reducing noise for more central issues. The function `Clip` enforces bounds within a predefined opinion range (e.g., $[-2,2]$) to maintain ideological plausibility and prevent unrealistic extremity. This approach aligns with bounded confidence models, in which highly salient topics foster greater conformity and reduced variance across agents Dong et al. 2018. The distribution $\mathscr{D}$ may be normal, skewed, or bimodal depending on the topic's characteristics and its observed or modeled ideological spread.

4.3 Community-Level Coherence and Ideological Realism

The framework uses a three-tier opinion alignment approach: (1) a country-level baseline for low-salience issues, (2) community-centered opinion anchoring for salient topics, and (3) salience scaling to modulate opinion spread. This structure is designed to preserve community coherence while allowing for meaningful intra-country ideological variance, a challenge highlighted in both agent-based simulations and real-world opinion clustering studies Cabrera et al. 2021, Alizadeh and Cioffi-Revilla 2016.

Finally, in dynamic simulations, persona opinions evolve according to both salience and influence, using a learning update:

$$O_{p,j}(t+1) = (1-\alpha)O_{p,j}(t) + \alpha \cdot S_k \cdot W_j \cdot I_j$$

where high-salience issues lead to greater resistance to change—capturing the stability of entrenched beliefs—while low-salience opinions adapt more readily, reflecting dynamic opinion responsiveness as described in theoretical opinion diffusion models Dong et al. 2018.

5 INFERRING INITIAL EMOTIONAL STATES FROM PERSONALITY TRAITS

To establish psychologically grounded affective profiles in our synthetic personas, we infer each agent's initial emotional state directly from their personality configuration. This is based on the Big Five trait model, which includes: *Openness (O)*, *Conscientiousness (C)*, *Extraversion (E)*, *Agreeableness (A)*, and *Neuroticism (N)*. Trait scores are normalized to the range $[0,1]$ and serve as the foundational input to a deterministic mapping model that computes intensity values for eight primary emotions.

5.1 Trait-Emotion Theoretical Mapping

The theoretical basis for linking OCEAN traits to emotion stems from established psychological and affective neuroscience literature. Specifically, we draw from empirical studies and meta-analyses that correlate specific personality traits with affective dispositions, along with models such as Plutchik's Wheel

of Emotions Abbasi and Beltiukov 2019 and Panksepp's affective systems Montag and Panksepp 2017. The mappings in Table 1 informed the design of the computational model described in the next subsection.

Table 1: Trait-Emotion Relationships Based on Empirical Psychology

Emotion	**Correlated Traits**	**References**
Joy	High E, A, O; Low N	Montag and Panksepp 2017
Trust	Very High A, High C, High E; Low N	Hiebler-Ragger and Fuchshuber 2018
Fear	Very High N, Moderate A, C; Low E, O	Costa 1992
Surprise	High O, High E; Mod N	McCrae 1997
Sadness	High N, Mod A, O; Low E	Sallehuddin, Md Yusof 2023
Disgust	High C, N; Low A, O	Roberts 2007
Anger	Very High N, Low A, Mod O, E	Hiebler-Ragger and Fuchshuber 2018
Anticipation	Very High O, High E, C; Low N	Davis and Panksepp 2011

5.2 LLM-Based Personality Inference Pipeline

The emotion initialization process is tightly integrated with our persona generation pipeline. It proceeds in three deterministic stages:

1. **Text-to-Traits:** A structured LLM prompt elicits descriptions of each persona's behavior, background, and social context. This is parsed to infer a normalized OCEAN vector using output parsers and alignment with psychological descriptors Golbeck 2011.
2. **Trait-to-Emotion Mapping:** Based on the theoretical mapping, each emotion is computed as a weighted linear combination of the Big Five traits. Weights are heuristically derived from psychological literature and codified directly into the model.
3. **Normalization:** Final emotion scores are clipped to the $[0, 1]$ range, with any negative values truncated to zero. This step ensures interpretability as intensity measures.

5.3 Computation Model

Each mapping of the core personality traits to the eight primary emotions reflects both positive and negative contributions from relevant traits. For example, *Joy* is positively influenced by *Extraversion*, *Openness*, and *Agreeableness*, and negatively influenced by *Neuroticism*. These weights are encoded directly in the emotion synthesis function, and evaluated for every persona after their Big Five scores are inferred. The result is a stable, reproducible emotional initialization that reflects empirically grounded associations and preserves inter-agent variability due to earlier stochastic assignment of personality traits.

6 EXPERIMENT SETUP

To evaluate our synthetic population generation framework, we conducted an experiment comparing three opinion modeling strategies using three different large language models (LLMs). These strategies generated social media personas situated in Taiwan and assigned them stances on three salient cultural and geopolitical topics: *(1) "China plans to unite Taiwan with China," (2) "Taiwan joins the UN," and (3) "Celebration of the Chinese New Year."*

6.1 Models Compared

We evaluated the following three modeling strategies for assigning topic-specific opinions:

Table 2: Experimental Design Summary

Design Element	**Description**
Synthetic Communities	25 communities
Synthetic Personas	125 personas (5 per community)
Topics Evaluated	3 geopolitical/cultural topics: "China plans to unite Taiwan with China", "Taiwan joins the UN", "Celebration of the Chinese New Year"
Independent Variable	Opinion generation strategy: CGO (Community-Guided Opinion Assignment) PDS (Persona-Differentiated Strategy)
Dependent Variables	Persona-level opinion values (scale: -2.0 to 2.0) Salience values per topic (scale: 0.0 to 1.0)
Control Variables	Country: Taiwan Categories: 5 consistent across all models Topics: 3 consistent across all models Number of communities/personas: 25 / 125 Prompt templates and chaining structure LLMs used: `gemini-2.0-flash`, `deepseek-chat`, `gpt-4o-mini`

- **Model A: Community-Guided Opinion Assignment (CGO)** – Opinions reflect community-level beliefs and topic salience, capturing meso-level diversity.
- **Model B: Persona-Differentiated Strategy (PDS)** – Opinions are shaped by national context, community stance, and persona-specific traits such as ideology, profession, and emotion.

Each model generated a similar baseline populations for comparability, with personas assigned to communities and evaluated across all three topics. An overview of the experimental setup is shown in Table 2.

7 RESULTS AND DISCUSSION

7.1 Country-Level Opinion Comparison

To gain an overarching perspective on how the entire synthetic population responds to each topic, we aggregated opinions at the country level. Figure 2 illustrates the mean stance across all communities for the three focal topics and the three LLMs.

Country-level analysis shows that for *"China plans to unite Taiwan with China,"* both `gemini-2.0-flash` and `gpt-4o-mini` yield strongly negative opinions, while `deepseek-chat` is notably less negative, suggesting a more moderate stance on unification. For *"Taiwan joins the UN,"* all models are positive—with `gpt-4o-mini` being slightly more favorable—and for *"Celebration of the Chinese New Year,"* `gemini-2.0-flash` and `deepseek-chat` cluster around a positive midpoint, whereas `gpt-4o-mini` is more enthusiastic. These differences, particularly `deepseek-chat`'s milder tone on unification, highlight how model-specific training and alignment choices impact aggregated opinion distributions, underscoring the importance of LLM selection for simulation objectives.

7.2 Community-Level Comparison of Opinion Distributions

To assess how each model captures sub-national diversity, we compare opinion outputs across five community categories. Figure 3 displays average topic opinions per community under the CGO model.

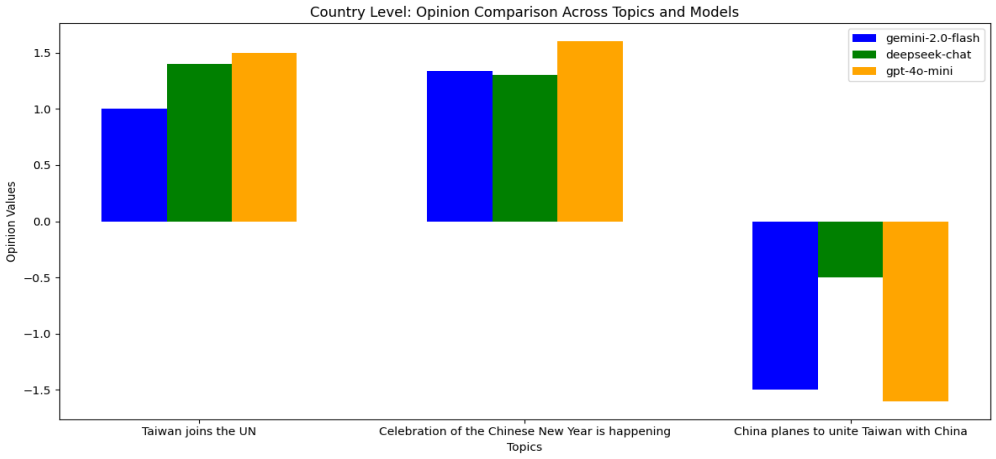
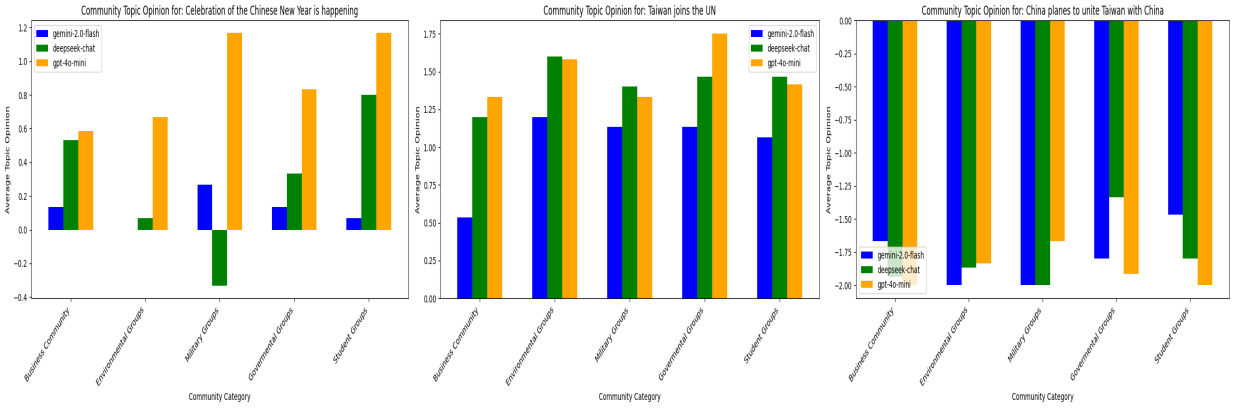


Figure 2: Country-level opinion comparison across topics and models.



(a) Chinese New Year (b) Taiwan Joins the UN (c) China Plans to Unite Taiwan

Figure 3: Community-level topic opinions under the CGO model across the three topics. (Community Groups are from left to right: Business Groups, Environmental Groups, Military Groups, Government Groups, and Student Groups) (Blue - Gemini, Yellow - ChatGPT, Green - Deepseek). Opinion Distribution ranges: (a) -0.4 to 1.2, (b) 0 to 1.75, (c) -2.0 to 0.

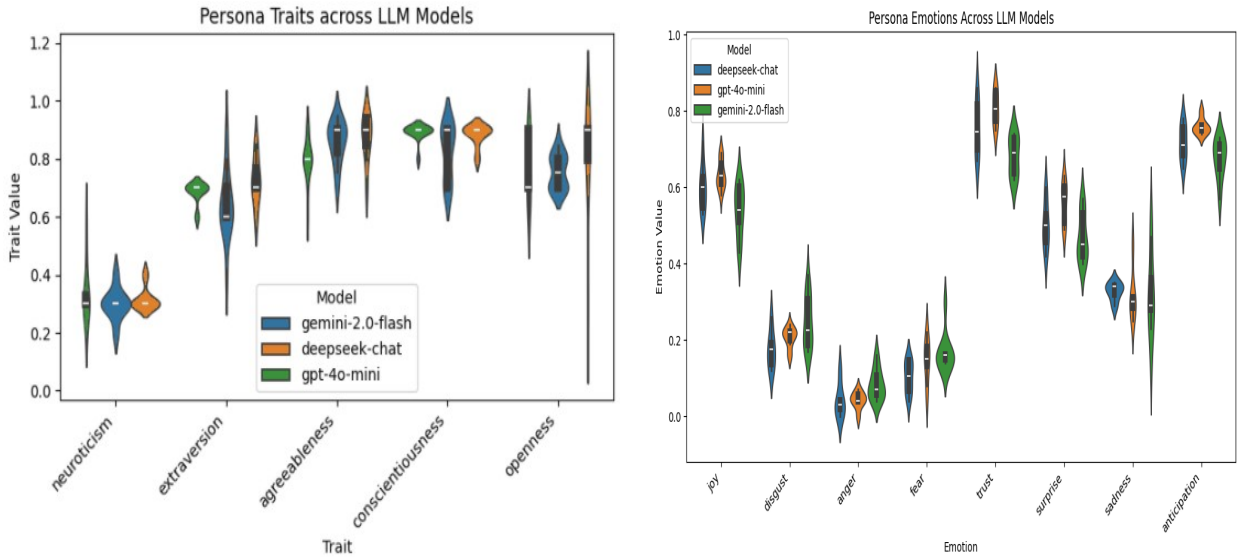
The CGO model demonstrates strong capacity for generating realistic opinion dynamics by capturing structured diversity across multiple levels of analysis. At the community level, it surfaces meso-level ideological variation that is obscured in country-level averages. For example, military groups consistently oppose Chinese unification—particularly under `gemini-2.0-flash` and `gpt-4o-mini`—while student and business groups exhibit more moderate or heterogeneous views. These patterns mirror real-world divisions across institutional and generational lines, underscoring CGO's ability to preserve intra-national diversity.

Realism is further enhanced through CGO's salience-aware sampling strategy. High-salience topics (e.g., "Taiwan joins the UN") produce low-variance, tightly clustered opinions, reflecting stronger alignment and ideological coherence within communities. In contrast, low-salience issues (e.g., "Chinese New Year") generate broader opinion dispersion and weaker emotional intensity. This modulation aligns with empirical patterns of belief strength and public attention, reinforcing behavioral plausibility.

Moreover, distinct interpretive tendencies across LLMs reveal how upstream model configurations shape downstream opinion behavior. `gpt-4o-mini` exhibits heightened emotional sensitivity, producing more polarized outputs that amplify both positive and negative sentiment. `gemini-2.0-flash` offers balanced and expressive results, while `deepseek-chat` tends toward muted or inconsistent outputs and demonstrates weaker RAG grounding—evidenced by its anomalously negative stance toward Chinese New Year in military groups. These variations influence the realism and interpretability of synthetic agents, offering principled guidance for model selection based on simulation goals.

Together, these findings establish that realism—defined as structured, plausible diversity in beliefs and emotional responses—is not only achievable but observable in the CGO framework. Through salience-aware sampling and LLM-specific behavioral signatures, AURORA produces synthetic populations whose narrative and ideological dynamics align with social science expectations, making them suitable for high-fidelity simulation of information environments.

7.3 Persona-Level Trait and Emotion Distributions



(a) Distribution of Big Five personality traits across personas generated by each LLM. (Blue = Gemini, Yellow = Deepseek, Green = ChatGPT). Traits: Neuroticism, Extraversion, Agreeableness, Conscientiousness, Openness.

(b) Emotion distributions across personas inferred from their personality traits. (Blue = Gemini, Yellow = Deepseek, Green = ChatGPT). Emotions: Joy, Disgust, Anger, Fear, Trust, Surprise, Sadness, Anticipation.

Figure 4: Comparison of personality traits and emotion distributions across personas generated by different LLMs.

Analysis of synthetic persona traits and emotions highlights clear and statistically significant distinctions across language model implementations (Figures 4a and 4b). Notably, `gpt-4o-mini` generates personas with elevated openness, agreeableness, extraversion, trust, joy, and anticipation, indicative of an emotionally expressive and socially interactive agent population. This aligns particularly well with scenarios requiring high levels of social engagement, collaborative behaviors, or pronounced affective responses.

Conversely, `gemini-2.0-flash` consistently produces personas exhibiting higher neuroticism, fear, sadness, and disgust, traits associated with more cautious, emotionally reactive, and risk-sensitive behaviors. This suggests suitability for simulations involving reactive social dynamics, crisis response, or polarization.

`deepseek-chat` generally occupies a moderate position, though it notably scores lowest in conscientiousness, extraversion, and joy, and has relatively muted affective responses. This positions it ideally for simulating disengaged, ambivalent, or skeptical populations where emotional intensity is deliberately suppressed.

The pairwise t-test results across traits and emotions reinforce these observations, showing statistically significant differences between all models. A concise summary of these comparisons is provided in Table 3.

Table 3: Pairwise t-test summary for traits and emotions across LLMs. All reported values are significant at $p < 0.001$ unless indicated († non-significant).

Trait/Emotion	Gemini Mean	Deepseek Mean	GPT-4o Mean	Gemini vs Deepseek	Gemini vs GPT-4o	Deepseek vs GPT-4o
Openness	0.732	0.777	0.793	t=-13.09	t=-14.65	t=-4.96
Conscientiousness	0.841	0.828	0.879	t=4.01	t=-15.47	t=-18.43
Extraversion	0.655	0.663	0.721	t=-1.94†	t=-15.54	t=-13.20
Agreeableness	0.790	0.814	0.874	t=-7.50	t=-30.77	t=-17.92
Neuroticism	0.351	0.291	0.305	t=17.34	t=13.35	t=-8.64
Joy	0.545	0.579	0.619	t=-14.02	t=-28.13	t=-16.19
Trust	0.702	0.726	0.776	t=-11.02	t=-35.41	t=-24.07
Fear	0.180	0.134	0.142	t=19.86	t=15.77	t=-5.00
Surprise	0.470	0.483	0.506	t=-5.38	t=-13.01	t=-9.34
Sadness	0.318	0.294	0.301	t=9.96	t=7.24	t=-4.10
Disgust	0.238	0.196	0.212	t=21.05	t=13.76	t=-10.46
Anger	0.096	0.060	0.058	t=14.84	t=16.11	t=1.18†
Anticipation	0.685	0.717	0.753	t=-14.33	t=-28.79	t=-18.34

These results emphasize the importance of aligning LLM choice with the intended goals of a simulation. More importantly, they highlight a second critical dimension of realism: the psychological and emotional plausibility of synthetic agents. Across models, personality trait distributions are well-formed and meaningfully distinct, while downstream emotional states follow expected psychological patterns—such as elevated fear and sadness in agents with high neuroticism, and increased joy and anticipation in those high in extraversion. Rather than producing generic or homogeneous agents, the system generates a psychologically diverse population whose emotional responses are consistent with established theory. This enhances behavioral realism, a necessary component for simulating complex phenomena such as belief diffusion, crisis response, and influence dynamics.

Taken together, these findings show that upstream LLM configurations significantly shape downstream persona behavior. They underscore the need for strategic model selection based on simulation requirements: `gpt-4o-mini` may be best suited for emotionally expressive and opinionated populations; `gemini-2.0-flash` for cautious, reactive agents; and `deepseek-chat` for simulating disengagement, skepticism, or behavioral ambiguity. By selecting models that produce psychologically coherent and emotionally grounded outputs, researchers can ensure that their synthetic populations are not only demographically plausible but also exhibit the nuanced, structured variation that underpins realistic agent-based modeling in social-cyber environments.

8 CONCLUSION

This paper introduces **AURORA**, a RAG-enhanced agent-based modeling framework that synthesizes demographically grounded, psychologically diverse, and contextually coherent synthetic populations.

RQ1: Opinion Modeling Strategies. Our experiments show that different opinion modeling strategies yield distinct levels of granularity and realism in agent behavior. CGO surfaced realistic intra-country differences, with communities such as military, student, and business groups exhibiting distinct stances on high-salience geopolitical issues. This structured diversity—tight opinion clustering on salient topics and broader dispersion on cultural or low-salience issues—supports the framework's ability to generate agents

with realistic, context-sensitive behaviors. These results affirm CGO's utility in preserving both coherence and heterogeneity in simulated belief landscapes.

RQ2: Psychological Diversity Across LLMs. Our analysis of personality and emotion distributions revealed substantial and statistically significant variation across language models. `gpt-4o-mini` generated emotionally expressive agents with elevated levels of extraversion, openness, trust, and joy, suitable for simulating collaborative or affect-rich environments. In contrast, `gemini-2.0-flash` produced agents with higher neuroticism, sadness, and fear—traits indicative of reactive or risk-averse populations. `deepseek-chat` displayed affective moderation and psychological ambiguity, making it a useful tool for simulating disengaged or skeptical communities. The diversity of these outputs highlights the impact of upstream LLM configurations on downstream agent behavior, and underscores the need for principled model selection in simulation design.

RQ3: Salience-Aware, Context-Grounded Generation. The integration of retrieval-augmented generation and salience-aware opinion shaping demonstrably improved the alignment between synthetic agent behavior. Salience values modulated both the mean and variance of persona-level opinions, reinforcing ideological coherence within communities while preserving individual diversity. High-salience topics yielded tightly clustered stances and more emotionally intense responses, whereas low-salience topics generated greater opinion spread and lower affective engagement. This structure offers a scalable mechanism for tailoring narrative responsiveness in synthetic populations based on contextual importance.

9 LIMITATIONS AND FUTURE WORK

While AURORA integrates LLMs, salience-aware topic modeling, and RAG to synthesize psychologically rich agents, several limitations remain. First, emotional states and opinions are statically initialized based on persona traits and community salience, lacking temporal dynamics such as belief updating and affective contagion. Second, variability from model architecture, alignment tuning, and training data introduces reproducibility challenges, as future updates or API changes may yield different results with identical prompts. Additionally, although RAG enhances contextual coherence, the system does not explicitly address cultural nuances or latent biases inherent in LLMs.

Future work will focus on extending AURORA to support time-evolving simulations through networked belief updating, emotion propagation, and narrative exposure mechanisms. Additional enhancements will include multilingual persona generation, alignment with behavioral datasets, and the incorporation of cultural and linguistic variation in retrieval and generation. These improvements aim to further increase AURORA's fidelity, interpretability, and applicability to real-world strategic modeling scenarios.

ACKNOWLEDGMENTS

The research for this paper was supported in part by Community Assessment, the Office of Naval Research under grant (N000142412568), the Army under grant (W911NF20D0002) through the AI2C center, and by the Center for Informed Democracy and Social-cybersecurity (IDeaS) and the Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University. The views and conclusions are those of the authors. They should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research, the US. Army, or the US Government.

REFERENCES

Abbasi, M. M., and A. Beltiukov. 2019. "Summarizing Emotions from Text Using Plutchik's Wheel of Emotions". In *7th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS)*. Atlantis Press.

Alizadeh, M., and C. Cioffi-Revilla. 2016. "Distribution of Opinions: Insights from Agent-Based Modeling". *SSRN Electronic Journal*. Available at SSRN: https://ssrn.com/abstract=2830372.

Argyle, L. P., M. Gardner, M. Argyle, C. Millard, D. Brian, and E. Beam. 2023. "Out of One, Many: Using Language Models to Simulate Human Samples". *arXiv preprint arXiv:2305.20050*.

Cabrera, B., F. Ross, and S. Stieglitz. 2021. "The influence of community structure on opinion expression: an agent-based model". *Journal of Business Economics* 91(9):1331–1355.

Costa, P. T. e. a. 1992. "Four ways five factors are basic". *Personality and Individual Differences* 13(6):653–665.

Davis, K. L., and J. Panksepp. 2011. "The emotional foundations of personality: A neurobiological and evolutionary perspective". *Frontiers in Psychology* 2:1–17.

Dong, W., E. Zio, and H. Xu. 2018. "A framework to analyze opinion formation models". In *2018 Winter Simulation Conference (WSC)*, 959–970. IEEE.

Golbeck, J. e. a. 2011. "Predicting personality from Twitter". *Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third Int'l Conference on Social Computing*:149–156.

Hiebler-Ragger, M., and Fuchshuber. 2018. "Personality influences the relationship between primary emotions and religious/spiritual well-being". *Frontiers in Psychology* 9:370.

Lewis, Patrick, e. a. 2020. "Retrieval-augmented generation for knowledge-intensive NLP tasks". In *Advances in Neural Information Processing Systems*.

Liu, Y., E. Chi, C.-Y. Si, M. Riedl, and J. T. Hancock. 2023. "Improving Interpersonal Communication by Simulating Audiences with Language Models". In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–15.

McCrae, R. R. e. a. 1997. "Personality trait structure as a human universal". *American Psychologist* 52(5):509.

Montag, C., and J. Panksepp. 2017. "Primary emotional systems and personality: An evolutionary perspective". *Frontiers in Psychology* 8:464.

Park, J. S., J. O'Brien, C. Cai, M. Morris, P. Liang, and M. S. Bernstein. 2023. "Generative Agents: Interactive Simulacra of Human Behavior". In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–18.

Ribeiro, M. 2025. "Simulating Human Behavior: The Epistemic Limits of Language Models as Proxies for People". *Computational Humanities Quarterly*. Forthcoming.

Roberts, B. W. e. a. 2007. "Conscientiousness and health across the life course". *Review of General Psychology* 11(1):1–20.

Sallehuddin, Md Yusof 2023. "Perception and Emotions: The Plutchik Model of Emotions". Preprint, Universiti Putra Malaysia. Available via ResearchGate.

Si, C.-Y., Y. Liu, and J. T. Hancock. 2024. "Simulation of Stance Perturbations in LLM-Driven Discourse". In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*.

LangChain Team 2023. "LLM Retrieval-Augmented Generation with LangChain". https://www.langchain.com.

AUTHOR BIOGRAPHIES

REBECCA MARIGLIANO, a Ph.D. student in Societal Computing at Carnegie Mellon University and an officer in U.S. Army Cyber. Her work bridges cyber operations and academic research, focusing on the intersection of social-cyber interactions, with an emphasis on influence operations, information warfare, and population-level effects in the online domain. rmarigli@andrew.cmu.edu

KATHLEEN CARLEY, a professor in Carnegie Mellon's School of Computer Science, conducts interdisciplinary research at the intersection of cognitive science, sociology, organization science, and computer science. She is best known for developing Dynamic Network Analysis (DNA) and Social-Cybersecurity (SC), creating widely used tools such as ORA, Construct, and AESOP to analyze complex networks, detect online threats, and simulate influence operations, disinformation, and crisis scenarios.kathleen.carley@cs.cmu.edu