twony: A Micro-Simulation of the Impact of OSN Mechanics on the Emotionality of Online Discourse

Simon Münker[1,*], Achim Rettinger[1,2]

[1] Trier University, Universitätsring 15, 54296 Trier, Germany
[2] FZI Research Center for Information Technology, Haid-und-Neu-Str. 10–14, 76131 Karlsruhe, Germany

Abstract

We introduce twony, a micro-simulation prototype designed to show the impact of online social network (OSN) mechanics — particularly recommendation algorithms — on emotional contagion and discourse dynamics. By harnessing the capabilities of Large Language Models (LLMs), the system generates an ecosystem of synthetic, politically engaged digital personas that interact within a controlled social media environment. These autonomous agents emulate human behaviors through in-context prompting techniques, enabling users to observe emotional transmission patterns under systematically varied conditions while circumventing the constraints inherent to real-user experimentation. The prototype implements two distinct recommendation paradigms: a baseline chronological feed and an emotion-prioritizing ranking mechanism that amplifies content based on emotional intensity, allowing the examination of the formation and reinforcement of echo chambers. Emotional valence is quantified via a fine-tuned BERT model, while network-level and agent-level visualizations track emotional cascades and polarization dynamics throughout the lifecycle. The system is architected as a browser-based application leveraging modern web technologies and decentralized APIs. Twony emphasizes accessibility, customizability, and extensibility. This contribution advances the field by providing a scalable, open-source prototype for systematically investigating OSN dynamics, offering actionable insights for platform designers and policymakers seeking to mitigate harmful emotional contagion while fostering healthier online discourse environments.

Keywords

Social Network Simulation, Recommendation Systems, Generative Agents, Language Models

1. Introduction

The proliferation of OSNs has significantly transformed the nature of digital discourse [1], enabling rapid information exchange while also amplifying emotional contagion [2]. As recommendation algorithms increasingly influence content visibility, concerns emerge regarding their role in shaping online discussions and reinforcing emotional polarization [3]. We developed twony (Fig. 1), a micro-simulation prototype, to explore the impact of OSN mechanics on the emotionality of digital interactions. By utilizing LLM-based agents, our system offers a controlled environment to demonstrate how different recommendation paradigms influence emotional contagion, discourse dynamics, and the emergence of echo chambers. Twony simulates a network of politically engaged digital personas interacting within a simplified social media ecosystem. The system implements two ranking mechanisms: a chronological feed and an emotion-prioritizing algorithm that amplifies content based on emotional intensity. We quantify emotional valence through a fine-tuned BERT classifier [4]. Unlike real-world studies, which face ethical and practical constraints in analyzing user behavior, twony is a live demonstration that mitigates privacy concerns while allowing for systematic manipulation of network variables. While the prototype serves as a demonstration tool rather than a scientific model of human behavior, its structured simulation of OSN interactions contributes to a better understanding of how algorithmic curation can shape emotional landscapes online. By offering a scalable and customizable platform,

SemGenAge: 1st Workshop on Semantic Generative Agents on the Web at ESWC 2025; Workshops and Tutorials Joint Proceedings
*Corresponding author.
muenker@uni-trier.de (S. Münker); rettinger@uni-trier.de (A. Rettinger)
http://simon-muenker.github.io (S. Münker); https://www.linkedin.com/in/achim-rettinger (A. Rettinger)
0000-0003-1850-5536 (S. Münker); 0000-0003-4950-1167 (A. Rettinger)

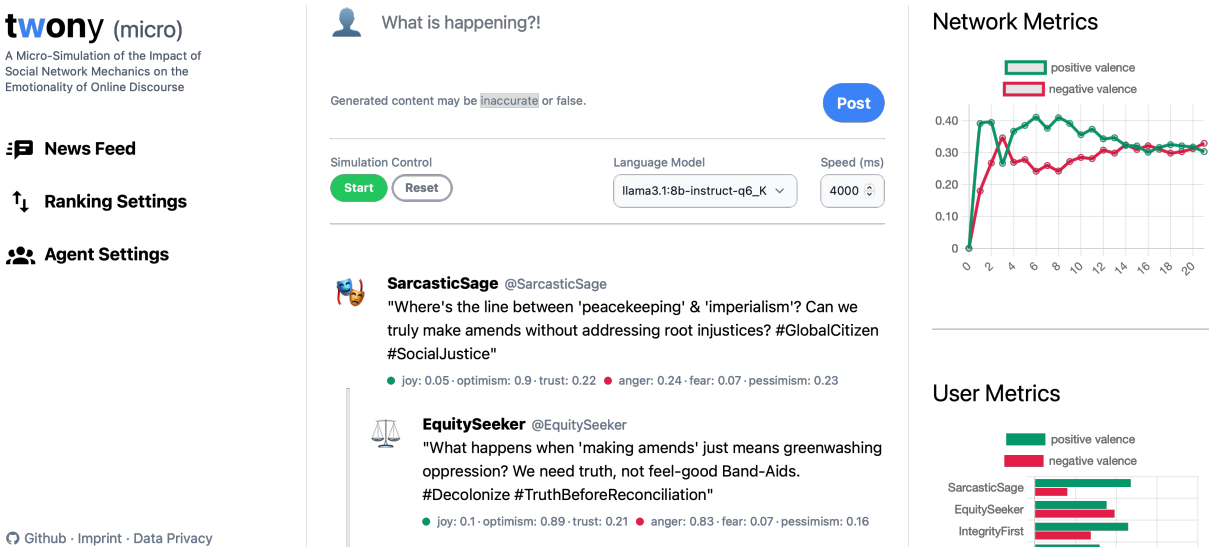

Figure 1: A screenshot of Twony during a simulated run: The interface features a three-panel layout inspired by the 𝕏 platform. The left column provides navigation, offering access to customization subpages and links to further information. The center column displays the actual social media feed and main control panel, allowing users to engage with agents and modify simulation parameters. The right column presents evaluation metrics tracked over time for each agent and includes functionality to download the application state.

twony provides insights for researchers, policymakers, and platform designers seeking to mitigate the adverse effects of emotional amplification while fostering healthier online discourse environments.

2. System Architecture

Our prototype is designed to deliver a fully autonomous generation of synthetic social media feeds by leveraging advanced LLMs to create and emulate a wide array of digital personas. These personas interact dynamically within a simulated social media environment, mimicking real-world online behaviors and discussions. While the system is capable of initiating and running simulations without any manual input, it also offers users the flexibility to intervene by manually posting content into the feed. This feature allows users to steer the direction of discussions toward specific topics of interest, enabling a more tailored and interactive experience. In the sections that follow, we provide a detailed overview of the specifications governing our agents, the architecture of the recommendation systems, and the methodologies employed for evaluation.

2.1. Agent Implementation (LLMs, Personalities, Interaction Patterns)

With recent advancements in generative AI and agent-based modeling, the key feature of our prototype is the adaptation of LLMs to specific personalities. Our system provides access to three state-of-the-art models: Llama 3.1 8B/70B [5], Mistral/Mixtral 7B/8x7B [6, 7], and Deepseek R1 7B/70B [8]. These models have different geographic origins: Llama was published by an American company, Mistral was developed in Europe, and Deepseek in China. This diversity enables us to test how discourses may differ between these LLMs and potentially reveal insights into their intrinsic biases [9, 10] resulting from training data selection and alignment processes. Furthermore, we provide both small and large versions of each model family to either facilitate rapid generation during live presentations or maximize the quality of generated content.

During the simulations, the selected model is adapted to different personas, authentically emulating individual social media users. We perform this alignment via in-context prompting, creating nuanced digital identities [11]. The description of each persona is extracted from a collection of 𝕏 tweets collected

throughout 2023, focusing specifically on users who actively engaged with political content—those who reacted to posts from politicians or media outlets. Our focus on politically engaged users aligns with twony's overarching mission to explore and understand democratic discourse in online debates. We automatically generate these comprehensive persona descriptions using LLMs to systematically analyze and summarize user behavior across multiple dimensions: *Humor Style, Communication Patterns, Emotional Expressions, Values and Beliefs, Interests and Hobbies, Social Interactions, Personality Traits, and Cultural Background.* The dimensions were partly selected based on preceding research on factors that influence online communication styles in synthetic agents [12]. This methodical approach ensures a consistently structured description for each persona while capturing their unique characteristics. Our flexible system allows for complete customization of these personas, while also supporting the seamless addition of new personas or removal of existing ones to adapt to evolving simulation needs (Fig. 2). We see the persona descriptions as a variable inside our system that the user should adapt to their use cases and our provided selection as demonstrative examples.

Our interaction mechanics employ a structured approach based on predetermined rules and probabilistic action models. The system evaluates each agent's likelihood of posting original content or replying to existing content at any given step. For replies, the system restricts candidate selections to the top-k ranked posts and implements a randomized selection process within this filtered pool. Additional constraints prevent agents from responding to their own uncommented posts, commenting on their immediately preceding comments, or creating consecutive original posts. With these constraints, we aim to maintain natural conversation flow and prevent artificial interaction patterns.

2.2. Network Topology and Recommendation Mechanics

We implement a fully connected network with a global shared feed. All agents receive the same ranked content which is displayed to users. The global feed can be sorted according to two distinct recommendation systems. As a baseline, we implement a chronological feed where the newest content appears at the top regardless of engagement metrics or emotional characteristics. Additionally, we define an emotion-based ranking that relies on the classification described in Section 2.3. The aggregated values — negative and positive valence — are combined to determine post-ranking. In the default settings, higher emotional intensity, regardless of valence type, yields a higher ranking. However, the system allows users to adjust the impact of these values within a range of ± 1 (Fig. 2). Consequently, both valence types can be configured to have either a negative impact on ranking or be neutralized entirely. With the emotion-based ranking system, we aim to model an echo-chamber effect for emotional content, hypothesizing that emotional intensity increases over time as agents are exposed to progressively more emotional content [13].

2.3. Evaluation Metrics (Emotion Classification)

For the evaluation, we utilize a pre-trained BERT classifier [4] trained to predict six emotions: joy, optimism, trust, anger, fear, and pessimism. We group these emotions into positive and negative valence categories [14] to determine their effect on users. This approach provides an opportunity to evaluate how discourse changes over time. In our demonstration, we display two aggregated views: an overarching network metric that shows emotional changes over time and a user-based metric that displays how each agent perceives and contributes to the network's emotional state. This dual visualization approach allows for tracking both macro-level emotional trends across the entire network and micro-level impacts of individual agents. The network-level visualization employs a time-series representation enabling the identification of significant emotional shifts and potential triggering events [15]. Thus, we can better understand how emotional contagion propagates through digital communities and potentially identify intervention points for mitigating negative emotional cascades.

3. Implementation Details

During the time of writing, the application is openly accessible at simon-muenker.github.io/TWONy-micro/, with the source code available in the GitHub repository github.com/simon-muenker/TWONy-micro. We leverage cutting-edge web technologies and decentralized architectures, emphasizing a high degree of customizability for end users and potential further adaptations.

Our prototype interface is implemented as a reactive JavaScript-based app that runs interactively as a browser application. The main system is served statically to the client and does not run server-side code. We utilize Astro as the website build engine in combination with Svelte [16] as the reactive UI framework. For designing our interface elements, we use the utility-first CSS framework Tailwind CSS [17]. For handling the local application state, we opt for the framework-agnostic store management system Nano Stores. We connect the LLMs and evaluation services via external APIs hosted separately on-premise. We deploy the LLMs via an Ollama backend through a customized API implemented in Python using FastAPI, exposing the necessary routes via a reversed NGINX proxy. The evaluation services are implemented in Python using the HuggingFace Transformers library [18] and FastAPI, also exposed through a reversed NGINX proxy. Through this decoupled architecture, we ensure streamlined adaptation and customization, such as replacing the provided LLM service with OpenAI or Anthropic interfaces.

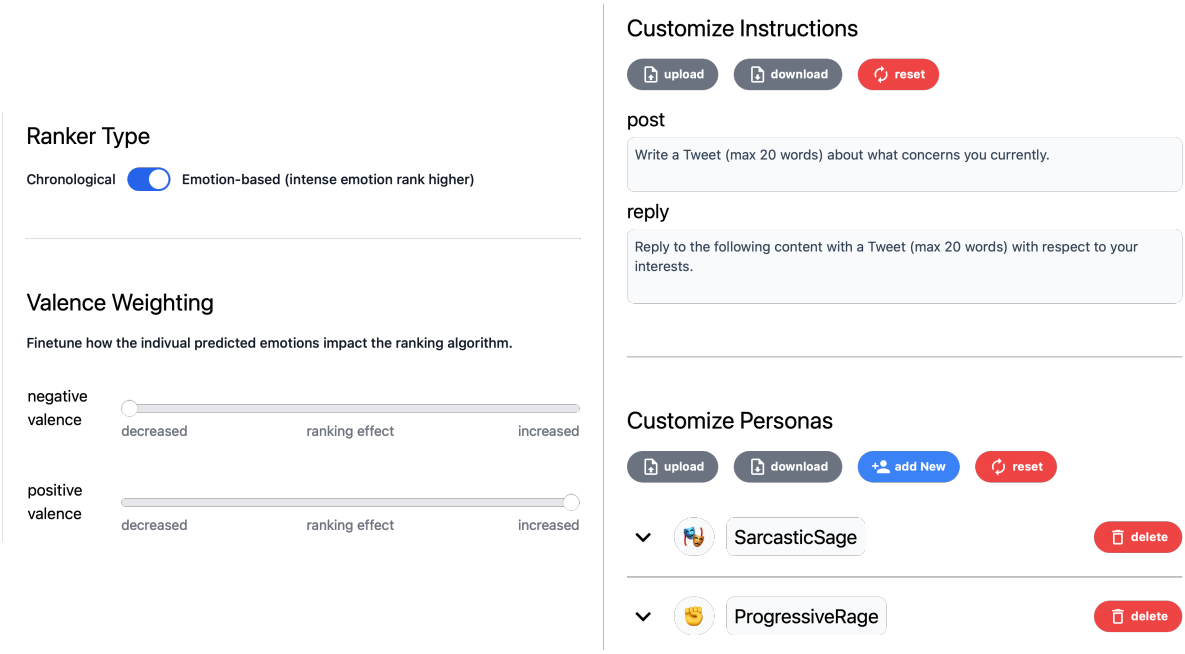

Figure 2: The twony ranking and agents settings page: Users can select their preferred recommendation system and adjust the weighting parameters that drive ranking algorithms Additionally, the platform offers control over agent configuration, allowing users to tailor textual instructions, modify persona descriptions, and personalize names and icons through via editing, replacing, deleting, or adding options.

4. Discussion

4.1. Implications for Understanding OSN Dynamics

The prototype demonstrates how recommendation algorithms that prioritize emotional content can potentially contribute to emotional contagion across digital platforms. By implementing both a neutral chronological feed and an emotion-prioritizing ranking mechanism, twony reveals how even simple algorithmic changes can significantly alter discourse patterns and emotional trajectories over time. Further, the simulation provides a controlled environment to observe the formation and reinforcement of

echo chambers. By tracking emotional polarization at both network and agent levels, twony illuminates how recommendation mechanics can inadvertently create feedback loops that amplify emotional intensity. The LLM-based agent architecture offers a novel perspective on user behavior modeling that bridges the gap between oversimplified theoretical models and ethically complex real-user experiments. By leveraging the sophisticated capabilities of LLMs to emulate human-like behaviors with consistent personas, the prototype provides a more nuanced representation of how diverse individuals might respond to and contribute to the emotional climate of online spaces.

4.2. Limitations of the Current Prototype

The current prototype employs a fully connected network with a global shared feed, which simplifies the complex topological structures observed in real OSNs. Actual social platforms feature clustered communities, varied connection strengths, and asymmetric influence patterns that significantly impact information flow [19] and emotional contagion. This simplified network topology may not adequately capture the nuanced dynamics of community formation and inter-group interactions that characterize real-world platforms. While the prototype LLM-based agents offer superficially convincing emulation of human-like behaviors, they remain imperfect approximations of actual user behavior. The personas, though derived from real social media data, cannot replicate the psychological complexity, contextual awareness, and historical experiences that shape human responses to social media content. Additionally, the current implementation lacks direct validation against observed human behaviors in comparable conditions. Also, our current focus on politically engaged users — who represent only a subset of typical social media users — creates an artificially engaged environment that doesn't reflect the true heterogeneity of OSNs. Most users are passive consumers or engage only occasionally with political content [20], creating different network dynamics than our simulation currently represents. Further, The emotion classification system, while functional, employs a simplified model of human emotions. The reduction to positive and negative valence categories, while methodologically sound, may obscure more nuanced emotional responses that influence discourse dynamics. Furthermore, emotional contagion in humans involves complex psychological mechanisms that may not be fully captured by the current simulation. Also, the prototype's current implementation does not account for external factors that significantly influence online discourse, such as breaking news events, seasonal trends, or platform-specific features like hashtags or groups. These contextual elements often serve as catalysts for emotional cascades in real OSNs and their absence may result in artificially stable or predictable simulation outcomes.

4.3. Potential Applications

Our prototype offers several promising applications across academic, and policy domains:

Policy Maker Lobbying Policymakers and regulators can leverage the prototype to assess in a simplified manner potential impacts of recommendation systems on digital platforms to the emotionality of political discourse. By adjusting parameters, twony allows for the visualization of how abstract changes might influence discourse patterns and emotional dynamics.

Digital Literacy Education The visual and interactive nature of the prototype makes it a suitable educational tool for demonstrating how recommendation systems influence information consumption and emotional responses. Educational institutions can incorporate the prototype in media literacy curricula to help students understand the mechanics behind their social media experiences and develop more critical engagement with digital platforms.

Computational Social Science Research While technically limited and designed as a demonstration tool, the prototype could provide researchers with a controlled environment to systematically study emotional contagion effects, polarization dynamics, and information diffusion patterns. This controlled testbed allows for isolating specific variables that would be difficult to manipulate in

studies with real users, potentially advancing theoretical understanding of online social dynamics. With the inclusion of LLMs from different geographic origins (American, European, and Chinese), the system offers an opportunity to examine how cultural contexts might influence online discourse patterns.

5. Conclusion

This work presents twony, a micro-simulation prototype designed to explore the impact of OSN mechanics on emotional discourse. By leveraging LLMs, twony simulates politically engaged digital personas interacting in a controlled social media environment. The prototype offers a systematic framework to analyze emotional contagion effects and the role of recommendation algorithms in shaping discourse. Additionally, prototype implements two ranking mechanism that contrasts chronological feeds with emotion-prioritizing ranking, allowing for a demonstration of how different content ranking paradigms influence the spread and polarization of emotional content. The system's open-source nature and modularity ensure extensibility for further research and policy evaluation, making it a valuable tool for studying digital discourse dynamics.

5.1. Key Insights

The prototype provides several insights into OSN dynamics. Most significantly, it showcases that recommendation algorithms that prioritize emotionally intense content can amplify emotional contagion and contribute to discourse polarization. It highlights how algorithmic curation plays a fundamental role in determining the tone and trajectory of online discussions. The prioritization of emotional intensity over time fosters self-reinforcing feedback loops, leading to the formation of echo chambers where exposure to diverse perspectives is minimized. This effect underscores the potential of digital platforms to shape ideological divides unintentionally. Furthermore, the use of LLM-based personas in twony exemplifies that agent-based simulation could offer a viable alternative to real-user studies, providing a controlled and ethical environment for studying digital discourse.

5.2. Improvements and Future work

While twony provides a first demonstration of emotional contagion in OSNs, several improvements and expansions can enhance its expressiveness. One key area for improvement is the selection of agent responses and post interactions. Future iterations should replace predefined interaction rules with more dynamic selection models that leverage LLM-driven decision-making or statistical approaches inspired by real-world behavioral data. Additionally, the current implementation of the prototype supports only open-weight LLMs available through Ollama. Expanding this selection to include closed-weight models such as GPT-4 [21], Claude, or Gemini [22] through API integration would improve the generative quality and enable broader comparative studies of digital discourse across different AI paradigms. Another critical enhancement involves refining the network topology. The current version of twony operates on a fully connected network, which does not accurately reflect the complexities of real-world OSN structures. Introducing community formations, asymmetric influence patterns, and varying connection strengths would better capture the nuances of online discourse [19]. Moreover, improving the emotional classification system by moving beyond a simple positive/negative valence model would provide a more accurate representation of emotional dynamics. Incorporating additional emotional categories and enabling real-time adaptation to shifts in discourse tone could significantly enhance the simulation's realism. A significant next step should involve validating the realism of agent behaviors before implementing more complex features. Drawing on methodologies from related work [23, 12], future work should compare our simulated behaviors against observed patterns in more sophisticated synthetic environments to ensure that our agents accurately reflect human behavior patterns. A further revelant step would be expanding the user base to include a more representative mix of personas, moving beyond politically engaged users to include the majority of network participants

who engage occasionally or not at all with political content. This will create more realistic network dynamics and better reflect the true heterogeneity of social media platforms. Finally, future work should consider integrating external contextual factors that influence emotional contagion in real OSNs, such as breaking news events, platform-specific trends, and evolving social movements. Incorporating these elements would increase the model's predictive power and make the simulations more applicable to real-world scenarios. While twony is primarily a demonstration tool, incorporating real-world user feedback and validation against empirical OSN data could further refine its utility for academic research and policy applications. By addressing these areas, twony can evolve into a more sophisticated and practical tool for studying, mitigating, and shaping healthier online discourse environments.

Acknowledgments

We thank Kai Kugler and Nils Schwager for the constructive discussions. This work is fully supported by twon (project number 101095095), a research project funded by the European Union under the Horizon framework (HORIZON-CL2-2022-DEMOCRACY-01-07).

Declaration on Generative AI

During the preparation of this work, the authors used Claude 3.7 and Grammarly in order to: Grammar and spelling check. After using these services, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] J. A. Tucker, Y. Theocharis, M. E. Roberts, P. Barberá, From liberation to turmoil: Social media and democracy, Journal of democracy 28 (2017) 46–59.

[2] A. Goldenberg, J. J. Gross, Digital emotion contagion, Trends in cognitive sciences 24 (2020) 316–328.

[3] J. Cho, S. Ahmed, M. Hilbert, B. Liu, J. Luu, Do search algorithms endanger democracy? an experimental investigation of algorithm effects on political polarization, Journal of broadcasting & Electronic media 64 (2020) 150–172.

[4] F. Barbieri, J. Camacho-Collados, L. Espinosa-Anke, L. Neves, TweetEval:Unified Benchmark and Comparative Evaluation for Tweet Classification, in: Proceedings of Findings of EMNLP, 2020.

[5] A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Yang, A. Fan, et al., The llama 3 herd of models, arXiv preprint arXiv:2407.21783 (2024).

[6] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M.-A. Lachaux, P. Stock, T. L. Scao, T. Lavril, T. Wang, T. Lacroix, W. E. Sayed, Mistral 7b, arXiv preprint arXiv:2310.06825 (2023).

[7] A. Q. Jiang, A. Sablayrolles, A. Roux, A. Mensch, B. Savary, C. Bamford, D. S. Chaplot, D. d. l. Casas, E. B. Hanna, F. Bressand, et al., Mixtral of experts, arXiv preprint arXiv:2401.04088 (2024).

[8] D. Guo, D. Yang, H. Zhang, J. Song, R. Zhang, R. Xu, Q. Zhu, S. Ma, P. Wang, X. Bi, et al., Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, arXiv preprint arXiv:2501.12948 (2025).

[9] A. Abid, M. Farooqi, J. Zou, Persistent anti-muslim bias in large language models, in: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, 2021, pp. 298–306.

[10] D. Rozado, The political biases of chatgpt, Social Sciences 12 (2023) 148.

[11] L. P. Argyle, E. C. Busby, N. Fulda, J. R. Gubler, C. Rytting, D. Wingate, Out of one, many: Using language models to simulate human samples, Political Analysis 31 (2023) 337–351.

[12] G. Rossetti, M. Stella, R. Cazabet, K. Abramski, E. Cau, S. Citraro, A. Failla, R. Improta, V. Morini, V. Pansanella, Y social: an llm-powered social media digital twin, arXiv preprint arXiv:2408.00818 (2024).

[13] M. Del Vicario, G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, W. Quattrociocchi, Echo chambers: Emotional contagion and group polarization on facebook, Scientific reports 6 (2016) 37825.
[14] L. F. Barrett, J. A. Russell, The structure of current affect: Controversies and emerging consensus, Current directions in psychological science 8 (1999) 10–14.
[15] N. Chetty, S. Alathur, Trigger event and hate content: Insights from twitter analytics, in: 2019 International Conference on Advances in Computing and Communication Engineering (ICACCE), IEEE, 2019, pp. 1–5.
[16] S. Bhardwaz, R. Godha, Svelte. js: The most loved framework today, in: 2023 2nd International Conference for Innovation in Technology (INOCON), IEEE, 2023, pp. 1–7.
[17] M. C. Klimm, Design Systems for Micro Frontends-An Investigation into the Development of Framework-Agnostic Design Systems using Svelte and Tailwind CSS, Ph.D. thesis, Hochschulbibliothek der Technischen Hochschule Köln, 2021.
[18] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, et al., Transformers: State-of-the-art natural language processing, in: Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations, 2020, pp. 38–45.
[19] P. Doreian, N. Conti, Social context, spatial structure and social network structure, Social networks 34 (2012) 32–46.
[20] W. Gong, E.-P. Lim, F. Zhu, Characterizing silent users in social media communities, in: Proceedings of the International AAAI Conference on Web and Social Media, volume 9, 2015, pp. 140–149.
[21] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, et al., Gpt-4 technical report, arXiv preprint arXiv:2303.08774 (2023).
[22] G. Team, R. Anil, S. Borgeaud, J.-B. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A. M. Dai, A. Hauth, K. Millican, et al., Gemini: a family of highly capable multimodal models, arXiv preprint arXiv:2312.11805 (2023).
[23] P. Törnberg, D. Valeeva, J. Uitermark, C. Bail, Simulating social media using large language models to evaluate alternative news feed algorithms, arXiv preprint arXiv:2310.05984 (2023).