GENEYE: AI-Driven Social Optics for Positive Feed

Kartik Pandey[a,*], Dr.J P Patra[b] Prabhudayal Vaishnav[c] and Priyanshu Tiwari [d]
[a] UTD-Chhattisgarh Swami Vivekanand Technical University, Bhilai.
[b] UTD-Chhattisgarh Swami Vivekanand Technical University, Bhilai.
[c] UTD-Chhattisgarh Swami Vivekanand Technical University, Bhilai.
[*] Corresponding Email: kartik0pandey00@gmail.com

Abstract

The exponential growth of social media platforms has intensified users' exposure to toxic language, misinformation, political polarization, and psychologically harmful content. Existing moderation systems are predominantly platform-centric, lack transparency, and provide limited personalization. This paper presents **GENEYE**, an AI-driven browser extension designed to enable real-time, user-controlled filtering of social media feeds.

GENEYE integrates a hybrid Natural Language Processing (NLP) framework that combines lightweight heuristic filtering, transformer-based classifiers, and Large Language Model (LLM) semantic scoring via OpenRouter APIs. The system applies a configurable rule-based decision engine that enforces user-defined thresholds to hide or blur undesirable content categories such as toxicity, hate speech, political discourse, sarcasm, rumors, and spam. Emphasis is placed on privacy preservation by maintaining all user preferences locally and minimizing external data transmission.

This paper describes the complete system architecture, AI/NLP workflow, research methodology, implementation details, and an evaluation framework focusing on accuracy, latency, cost efficiency, and user experience. The proposed approach demonstrates the feasibility of deploying intelligent, transparent, and personalized content moderation directly at the browser level.

Keywords— Content Moderation, Social Media Filtering, Natural Language Processing, Large Language Models, Browser Extensions, Human-in-the-Loop AI, Digital Well-being.

References

[1] vaswani2017attention Vaswani, A., et al. "Attention is all you need." NeurIPS, 2017.

[2] Devlin, J., et al. "BERT: Pre-training of deep bidirectional transformers." NAACL, 2019.

[3] OpenAI. "GPT-4 Technical Report." arXiv preprint, 2023.

[4] Schmidt, A., Wiegand, M. "A survey on hate speech detection." ACL, 2017.

[5] Google. "Chrome Extensions Manifest V3 Documentation."

[0]International Conference on Mathematical Sciences and Computational Intelligence (ICMSCI-2026).