LISTEN: Leveraging Open-Weight LLMs for Real-Time Monitoring of Health Misinformation on Social Media

Background

Misinformation about transmission, symptoms, and treatment has repeatedly fueled confusion and stigma during disease outbreaks. While prior efforts such as WHO's EARS platform provided infodemic monitoring, they were limited in scope and longevity [1,2]. LISTEN (Leveraging Intelligence from Social Networks to Promote Health Equity and Neutralize Negative Content) addresses these gaps by offering a transparent, adaptable, and stakeholder-informed system for real-time analysis of online discourse.

Methods

We implemented LISTEN in Python/Streamlit with open-weight LLMs. The system collects weekly posts from X using keyword filters across multiple disease topics (e.g., "cancer," "measles," "mpox," "monkeypox," "mpxv"), normalizes metadata (engagement, timestamp, verification), and classifies misinformation into four themes: *symptoms*, *transmission*, *treatment*, and *other*. Classification is performed via few-shot in-context learning. A subset of posts was manually annotated to establish ground truth, guided by prior literature [3,4] and refined in collaboration with public health experts. We validated model performance using precision, recall, accuracy, and macro-F1, and collected feedback from potential stakeholders.

Results

LISTEN curated over 21,000 unique posts published between July 2022 and June 2025. Key opinion leaders included journalists, health professionals, and influencers, underscoring the importance of monitoring high-visibility accounts in shaping discourse. The misinformation detection module achieved a macro-F1 score of 0.806 (precision 0.556, recall 0.833, accuracy 0.907) with 17-shot prompting, substantially outperforming zero-shot baselines. Stakeholder feedback emphasized the utility of engagement visualizations, rationale transparency, and the potential for integration with bot detection and counter-messaging features.

Conclusions

The LISTEN prototype [5] demonstrates the feasibility of integrating interpretable, open-weight LLMs into real-time public health surveillance. By surfacing emerging misinformation narratives and their amplification pathways, the system enables evidence-informed communication strategies to reduce stigma and misinformation harms. Future work includes developing counter-messaging interventions, extending to additional diseases and platforms, and evaluating impacts on public understanding and behavior.

[1] Eysenbach, G., et al. (2020). How to fight an infodemic: the four pillars of infodemic management. Journal of Medical Internet Research, 22(6): e21820.
[2] Yan, X., Li, Z., Cao, C., et al. (2024). Characteristics, influence, prevention, and control measures of the mpox infodemic: Scoping review of infodemiology studies. JMIR, 26: e54874.
[3] McGowan, B. S. (2022). World Health Organization's early AI-supported response with social listening platform. Journal of the Medical Library Association (JMLA), 110(2): 273.
[4] Du, M., Qin, C., Yan, W., et al. (2023). Trends in online search activity and the correlation with daily new cases of monkeypox among 102 countries or territories. Int'J Envir Research & Public Health, 20(4): 3395.
[5] Tang, L. Y. W. (2025). LISTEN: a prototype. Available at: https://measles.streamlit.app/