

WHEN AI JOINS THE CROWD

Understanding & Designing AI Influence In Mixed Human-AI Online Environments

Nethmi Wijesinghe[1,2], Dr Danula Hettiachchi[1,2], Dr Damiano Spina[1,2], Dr Senuri Wijenayake[1]

[1]RMIT University, [2]ARC Centre of Excellence for Automated Decision Making + Society

Social Conformity | Human-AI Groups | Generative-AI | Decision Making | Responsible AI

Online platforms now integrate AI-generated responses and recommendations alongside human opinions, creating mixed human-AI environments. These opinions can encourage **social conformity**, prompting people to change their judgements to align with others. This influence may depend on **contextual (majority opinion and majority group size), personal (user's confidence in their own judgement), and AI-related (AI agent's position within the group and its opinion framing strategy) factors**. We examine how these factors jointly shape conformity and decision-making, and whether judgement changes improve or reduce accuracy. The goal is to inform responsible AI design that supports informed, independent judgement.

STUDY ONE - COMPLETED

Examining Social Conformity to Mixed Human-AI Opinions in News Trustworthiness Judgements

We aimed to understand how established conformity determinants, together with AI position and opinion framing shape social conformity in news trustworthiness judgements within mixed human-AI environments. We also examined whether these judgement changes improve or reduce accuracy and how they relate to subsequent engagement intentions.

Participants, Task and Stakeholder Involvement

Conducted crowdsourcing task with **120 regular social media users** recruited via Prolific.

Assessed 20 Facebook-style news posts (10 real and 10 fake) before and after viewing human and/or AI opinions. Rated each post's perceived trustworthiness, confidence in the judgement and the likelihood of engaging with the post.

Mixed Methods

Linear Mixed Effects (LMM) statistical models.
Paired t-tests.
Collaborative deductive thematic analysis.

Results

Does the crowd still matter?

Larger majority – minority gaps → stronger conformity

Critical majorities (judging posts as fake) showed influenced conformity than supportive majorities (judging posts as real).

Can AI reinforce or challenge the crowd?

AI supports the human majority → conformity increases
AI opposes the human majority → conformity decreases

Does AI framing matter?

Both authoritative and peer-style AI significantly influenced judgements, with authoritative framing producing stronger influence.

Can AI help or mislead?

Following incorrect AI despite a correct majority → accuracy decreases

Following correct AI despite an incorrect majority → accuracy increases

What happens to engagement?

With substantial trustworthiness changes, conforming responses showed increased reacting, commenting and sharing intentions under supportive majorities and decreases under critical majorities. Non-conforming responses showed the reverse.

STUDY TWO - ONGOING

Designing the AI Voice: Co-designing AI Opinion Framings for Mixed Human-AI Environments

We aim to co-design AI opinion framings strategies for mixed human-AI online learning environments. We also explore which design dimensions they prioritize and how these framings could apply to other online contexts.

Participants, Task and Stakeholder Involvement

Conduct 5 co-design workshops with **12 learners** with experience using an LMS and **8 educators** with experience teaching, designing or supporting learning through one.

Generate ideas individually and collaboratively develop AI opinion framings for mixed human-AI learning environments. Identify and prioritize key design features.

Qualitative Methods

Braun and Clarke's 6-phase reflexive thematic analysis approach.

FUTURE DIRECTIONS

Conduct a crowdsourcing study with online learners to examine how different AI opinion framing strategies shape conformity across objective and subjective tasks.

Conduct a lab study of AI errors over time to examine how conformity changes as users encounter repeated AI errors.

PATHWAY TO IMPACT

Translate findings into practical guidance for presenting AI opinions. Evaluate proposed framings with users and share findings through publications, presentations and accessible resources for designers and practitioners.

This research addresses the risks of misleading information and uncritical reliance on AI across online environments.

Post your recommendations here!

This research was conducted by the ARC Centre of Excellence for Automated Decision-Making and Society (CE200100005) and funded fully by the School of Computing Technologies Scholarship and International Tuition Fee Scholarship offered by the RMIT University.

Main image and icons are generated with GPT-6.1 Sol (Extra High)

References

1. Wijesinghe, N. (2026, July). Characterising the Effects of Social Conformity on Information Seeking in Mixed Human-AI Environments. In Proceedings of the 49th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 5318-5318).
2. Wijenayake, S., Hettiachchi, D., Hosio, S., Kostakos, V., & Goncalves, J. (2020). Effect of conformity on perceived trustworthiness of news in social media. IEEE Internet Computing, 25(1), 12-19.
3. de Jong, S., Jacobsen, R. M., Wester, J., Wijenayake, S., Goncalves, J., & van Berkel, N. (2025, June). Impact of Agent-Generated Rationales on Online Social Conformity. In Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (pp. 3370-3384).
4. Govers, J., Sew, C., Velloso, E., Kostakos, V., & Goncalves, J. (2026, April). Narratives and Perspectives: How AI Summaries Steer Users' Opinions and Engagement on Social Media. In Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (pp. 1-19).