

Rethinking AI-Mediated Minority Support in Power-Imbalanced Group Decision-Making: From Anonymity To Authenticity

Soothan Lee

Department of Design, UNIST
Ulsan, Republic of Korea
soohtwanlee@unist.ac.kr

Kyungho Lee

Department of Design, UNIST
Ulsan, Republic of Korea
kyungho@unist.ac.kr

Abstract

AIMC systems increasingly aim to protect minority voices by anonymizing or proxying their input, but anonymity and authenticity are not the same construct. This position paper draws on an ongoing empirical study comparing two LLM-powered minority support strategies in hierarchical group decision-making. We found that relaying minority input anonymously through AI increased participation but significantly reduced psychological safety and satisfaction, while generating only autonomous counterarguments improved satisfaction and reduced marginalization. These counterintuitive findings reveal three provocations for AIMC design in hierarchical contexts: the inherent trade-offs among anonymity, authenticity, agency, and accountability; the risk that power asymmetry reverses intended effects; and the need for AI to facilitate group reflection rather than substitute for human responsibility. These findings and provocations are offered as a contribution to the *Restoring Human Authenticity in AI-Mediated Communication* workshop.

CCS Concepts

• **Human-centered computing** → **Computer supported cooperative work**; *Collaborative interaction*; *Natural language interfaces*; *HCI theory, concepts and models*.

Keywords

Group-centered AI, AI-mediated Communications, Decision-making, Minority Influence, Group Dynamics

ACM Reference Format:

Soothan Lee and Kyungho Lee. 2026. Rethinking AI-Mediated Minority Support in Power-Imbalanced Group Decision-Making: From Anonymity To Authenticity. In *Workshop on Restoring Human Authenticity in AI-Mediated Communication (AI-MC) at the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 3 pages. <https://doi.org/XX.XXXX/XXXXXXXX.XXXXXXX>

1 Motivation and Background

Power imbalances in group decision-making frequently suppress minority perspectives and weaken collective outcomes [4]. Compliance pressures lead individuals to publicly align with the majority despite private disagreement, undermining psychological safety and discouraging participation from less-empowered members [5, 9]. Recent HCI research has introduced AI-powered interventions to

mitigate groupthink and encourage broader participation [1], but many assume relatively equal standing among participants. Comparatively less explored is how AI can represent and route dissent when dissenters face social risk from identification or retaliation in power-imbalanced situations [3, 6, 8].

This position paper draws on an ongoing empirical study comparing two LLM-powered minority support strategies in hierarchical group decision-making. We found that anonymously relaying minority input through AI increased participation but significantly reduced psychological safety and satisfaction, while generating only autonomous counterarguments improved satisfaction and reduced marginalization. These counterintuitive findings suggest that in hierarchical settings, how AI introduces dissent and whose voice it is perceived to carry matters as much as whether dissent enters the discussion at all. This paper contributes to *CHI'26 Workshop on Restoring Human Authenticity in AI-Mediated Communication* by arguing that anonymity and authenticity are distinct properties that current AIMC designs tend to conflate. We summarize our empirical findings, discuss how power asymmetry complicates assumptions in existing AIMC work, and offer design provocations for preserving expressive ownership in hierarchical group settings.

2 Summary of Empirical Findings

We conducted a mixed-methods experiment with 96 Korean participants in 24 four-person groups, each consisting of three high-power majority members (seniors) and one low-power minority member (junior). Groups completed two corporate decision-making tasks under one of three conditions (Figure 1): a no-AI baseline, an AI-generated Counterargument condition (AIGC), or an AI-mediated Message condition (AIMM). In AIGC, the system autonomously posts counterpoints to the emerging majority consensus. In AIMM, the system also accepts private input from the minority member, paraphrases it, and relays it (Figure 2-C) as its own output alongside its own counterarguments, making the two message types indistinguishable in terms of provenance. We measured psychological safety, marginalization, engagement, satisfaction, and cognitive workload via validated self-report scales and behavioral metrics, supplemented by semi-structured exit interviews.

Final group decisions aligned with senior preferences in 80% of cases across all conditions, suggesting that neither intervention was sufficient to shift collective outcomes. AIGC reduced minority members' marginalization and improved their satisfaction without negatively affecting psychological safety, as autonomous counterarguments fostered a more open group atmosphere in which minority members felt less isolated in their dissent. In contrast, AIMM increased participation volume but significantly reduced psychological safety and satisfaction. Majority members discounted

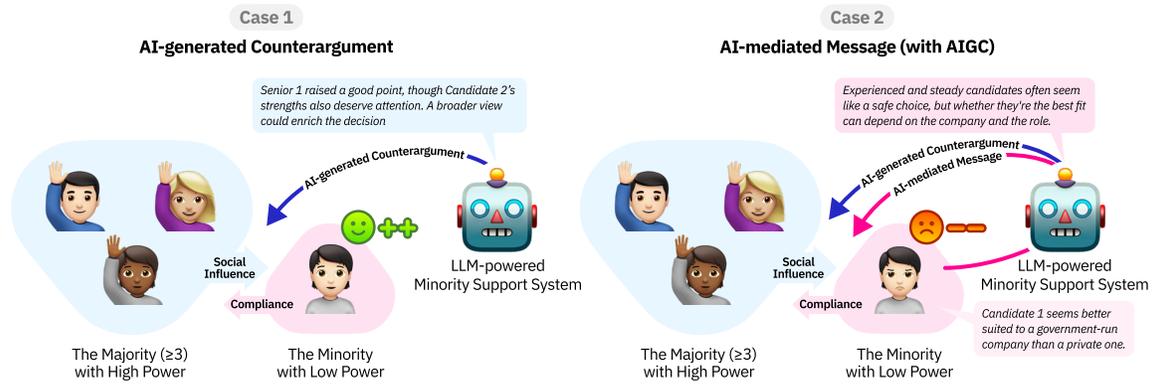


Figure 1: An LLM-powered minority support system mediates majority-minority dynamics through two designs: AIGC, which autonomously posts counterpoints to normalize dissent, and AIMM, which paraphrases privately submitted minority views blended with AIGC outputs to obscure authorship. AIGC improved group atmosphere and satisfaction, whereas AIMM increased participation but paradoxically undermined minority members' psychological safety and satisfaction.

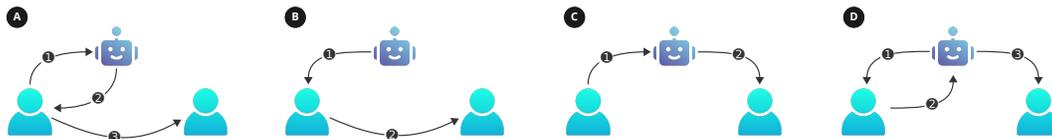


Figure 2: Four patterns of AI-mediated group communication[7]: (A) a human relays AI-generated content to the group; (B) a human selectively discloses AI output to participants; (C) the AI reformulates a human's message before presenting it to the group; and (D) the AI directly facilitates inter-participant communication. Arrows denote information flow, with numbered sequences indicating interaction order.

AI-authored messages because they perceived them as non-human, leaving minority members feeling ignored even when their views were relayed. This produced invisibility — authorship was hidden, and the message was devalued as AI-generated.

3 Future Design Provocations

Our findings suggest three tensions that AIMC research and design should address more directly in hierarchical group contexts.

Anonymity, Authenticity, Accountability, and Agency Involve Inherent Trade-offs. Anonymity is often discussed as one mechanism for protecting minority voices [10], but our results suggest it is neither sufficient nor straightforwardly beneficial in hierarchical settings. Full anonymity allows minority members to express dissent without social risk, but can simultaneously strip them of expressive ownership, visibility, and the sense that their contribution is genuinely theirs. The inverse is equally problematic: speaking directly preserves authenticity, agency, and accountability, but exposes the speaker to marginalization and psychological distress. Designing for authenticity in AIMC therefore requires attending not only to anonymity, but to how systems can support minority members in feeling genuinely heard, credited, and responsible for their own contributions.

Power Asymmetry Can Reverse the Intended Effects of AIMC. Most AIMC systems are designed and evaluated under conditions of relatively equal participation. In hierarchical settings,

however, AI-authored messages may be systematically discounted by dominant group members, who may actively rebut AI interventions to reinforce their own position rather than engaging with the underlying dissent. What is designed to foster inclusion can, in practice, deepen existing power imbalances. Future AIMC research should treat power asymmetry as a primary design variable rather than a contextual footnote.

AI Should Prompt Reflection, Not Replace Organizational Responsibility. The organizational contexts where such systems are most needed may also be least likely to adopt them, or may adopt them in ways that serve majority rather than minority interests. Effective deployment therefore requires examining not only system efficacy but also the adoption incentives of senior stakeholders and the motivations of members who will use the system. More broadly, future systems might more sustainably foster inclusive atmospheres by prompting collective reflection and encouraging groups to find better directions on their own. Ultimately, the goal should be organizations that no longer need such systems because psychological safety [2] is already embedded in their culture.

4 Conclusion

This position paper argues that anonymity and authenticity are distinct constructs that AIMC design must address separately. Our empirical findings show that AI-mediated anonymity can increase participation while simultaneously eroding expressive ownership

and psychological safety. Future AIMC systems should move beyond protecting the volume of minority speech toward ensuring that minority voices are genuinely heard, attributed, and acknowledged within their group's relational dynamics.

Acknowledgments

This research was partially supported by a grant from the Korea Institute for Advancement of Technology (KIAT) funded by the Government of Korea (MOTIE) (P0025495, Establishment of Infrastructure for Integrated Utilization of Design Industry Data).

References

- [1] Chun-Wei Chiang, Zhuoran Lu, Zhuoyan Li, and Ming Yin. 2024. Enhancing AI-Assisted Group Decision Making through LLM-Powered Devil's Advocate. In *Proceedings of the 29th International Conference on Intelligent User Interfaces (IUI '24)*. Association for Computing Machinery, New York, NY, USA, 103–119. doi:10.1145/3640543.3645199
- [2] Amy Edmondson. 1999. Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly* 44, 2 (June 1999), 350–383. doi:10.2307/2666999
- [3] Angel Hsing-Chi Hwang and Andrea Stevenson Won. 2024. The Sound of Support: Gendered Voice Agent as Support to Minority Teammates in Gender-Imbalanced Team. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–22. doi:10.1145/3613904.3642202
- [4] Irving L. (Irving Lester) Janis. 1982. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. Boston: Houghton Mifflin.
- [5] Herbert C. Kelman. 1958. Compliance, Identification, and Internalization Three Processes of Attitude Change. *Journal of Conflict Resolution* 2, 1 (March 1958), 51–60. doi:10.1177/002200275800200106
- [6] Soohwan Lee, Seoyeong Hwang, Dajung Kim, and Kyungho Lee. 2025. Conversational Agents as Catalysts for Critical Thinking: Challenging Social Influence in Group Decision-making. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3706599.3719792
- [7] Soohwan Lee, Mingyu Kim, Seoyeong Hwang, Dajung Kim, and Kyungho Lee. 2025. Amplifying Minority Voices: AI-Mediated Devil's Advocate System for Inclusive Group Decision-Making. In *Companion Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25 Companion)*. Association for Computing Machinery, New York, NY, USA, 17–21. doi:10.1145/3708557.3716334
- [8] Xiaoyan Li, Naomi Yamashita, Wen Duan, Yoshinari Shirai, and Susan R. Fussell. 2022. Improving Non-Native Speakers' Participation with an Automatic Agent in Multilingual Groups. *Proc. ACM Hum.-Comput. Interact.* 7, GROUP (Dec. 2022), 12:1–12:28. doi:10.1145/3567562
- [9] Serge Moscovici and Elisabeth Lage. 1976. Studies in Social Influence III: Majority versus Minority Influence in a Group. *European Journal of Social Psychology* 6, 2 (1976), 149–174. doi:10.1002/ejsp.2420060202
- [10] Dimitri Staufer, Frank Pallas, and Bettina Berendt. 2024. Silencing the Risk, Not the Whistle: A Semi-automated Text Sanitization Tool for Mitigating the Risk of Whistleblower Re-Identification. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 733–745. doi:10.1145/3630106.3658936