expressive computing lab.

expc

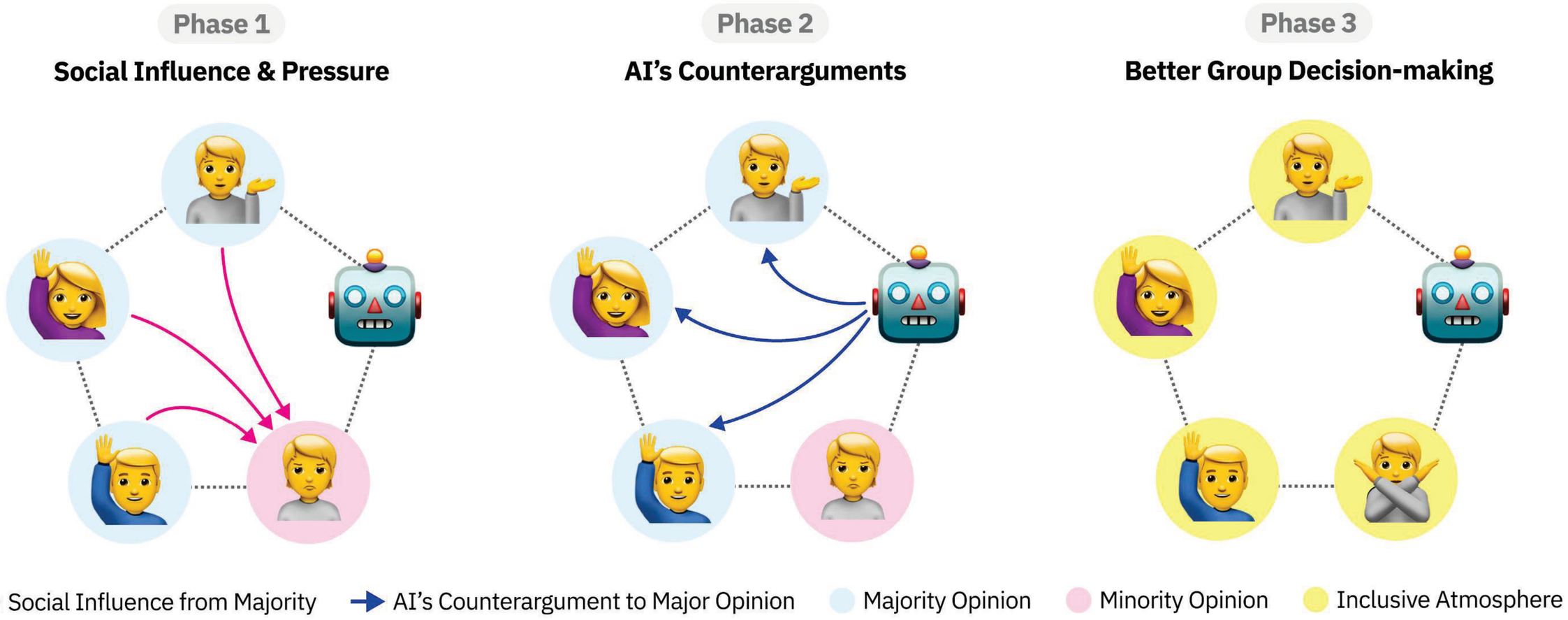
😈 Conversational Agents as Catalysts for Critical Thinking
Challenging Social Influence in Group Decision-making

*Soohwan Lee, *Seoyeong Hwang, Dajung Kim, Kyungho Lee
EXPC Lab, Department of Design, Ulsan National Institute of Science and Technology

PAPER WEB SITE

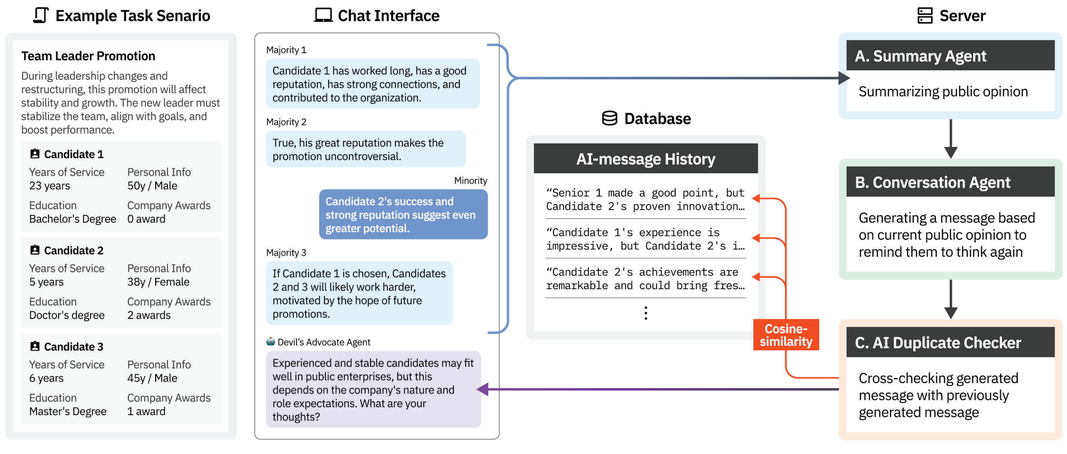
Overview

- 🔍 **Problem addressed**: This study tackles how social influence and power dynamics in groups can suppress minority viewpoints, leading to groupthink and poor decisions. Specifically, it examines situations where low-power members feel pressured to ***comply*** with majority opinions rather than expressing dissenting views.
- 💡 **Proposed solution**: The researchers developed an LLM-powered "Devil's Advocate" system that automatically generates counterarguments during group discussions. This AI agent is designed to challenge majority opinions, legitimize alternative viewpoints, and foster more inclusive decision-making environments.
- 📊 **Key findings**: The experimental results showed that AI-generated counterarguments fostered a more flexible atmosphere and significantly enhanced satisfaction with both the decision-making process and outcomes for all participants. Improvements were particularly notable for minority (junior) members, though there was a slight (non-significant) increase in cognitive workload.



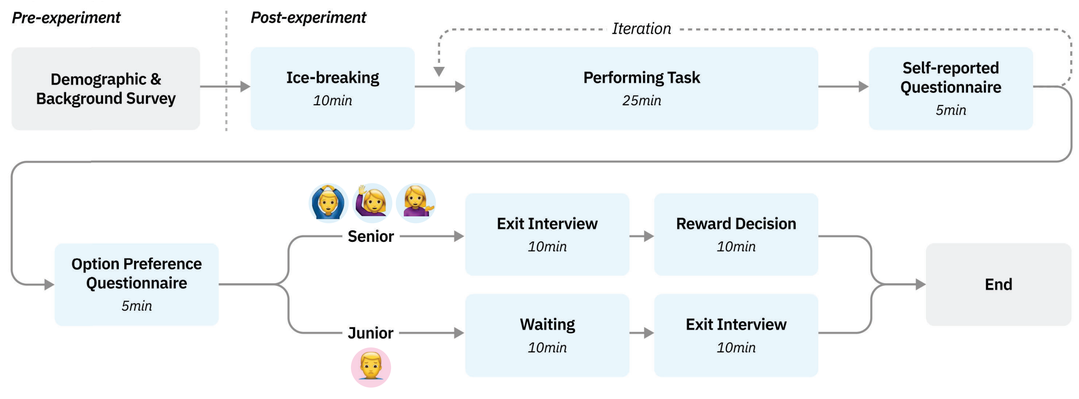
System Implementation

The system implementation features a real-time chat environment built with TypeScript (React) and Python (FastAPI), integrated with GPT-4o. The architecture includes three key components: a Summary Agent that consolidates emerging consensus, a Conversation Agent that generates empathetic counterarguments through Socratic questioning, and an AI Duplicate Checker that prevents repetitive content. The system intervenes after approximately eight human messages, ensuring balanced participation while maintaining discussion flow. This design employs empathetic communication styles, utilizes Socratic questioning to promote critical thinking, and facilitates anonymous communication to enhance psychological safety and prevent groupthink.



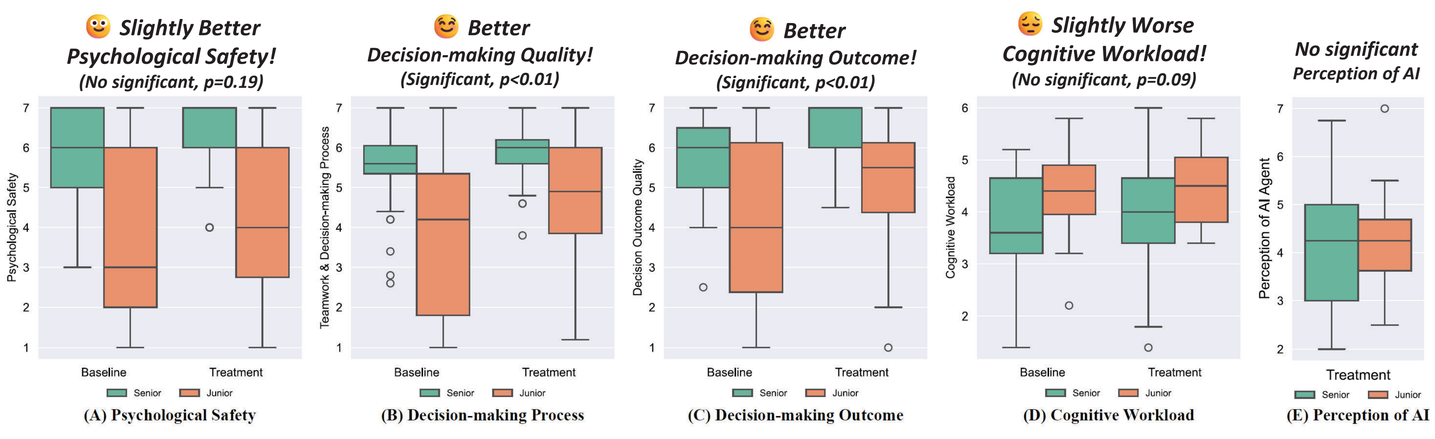
Experimental Design

The study involved 48 Korean participants (aged 19-39) organized into 12 groups of four, with each group containing *three high-power "senior" members* and *one low-power "junior" member*. Power dynamics were established through role titles and compensation structure, with seniors receiving higher base compensation and having discretion to award juniors additional rewards based on contribution assessment. The experimental design employed a mixed methodology, with Participant Type (senior/majority vs. junior/minority) as a between-subjects variable and Communication Condition as a within-subjects variable, allowing each participant to experience both baseline and AI intervention conditions. Participants completed two 20-minute corporate-context tasks (evaluating employee profiles for promotion and analyzing potential contract partners), designed to create natural majority-minority dynamics through situational context rather than explicit role-playing instructions. After each task, participants completed questionnaires measuring psychological safety, decision satisfaction, and cognitive load using 7-point Likert scales. Data analysis employed robust regression with mixed models followed by Tukey post-hoc tests to compare conditions and participant types.



Quantitative Results

The results showed psychological safety increased slightly but non-significantly in the treatment condition (from 5.38 to 5.65), with seniors consistently reporting higher scores than juniors. Satisfaction with both decision-making process and outcomes improved significantly (process: 5.10 to 5.55; outcomes: 5.31 to 5.89), benefiting both participant types though juniors consistently reported lower satisfaction than seniors. Cognitive workload increased slightly but non-significantly (3.93 to 4.12), with juniors experiencing higher workload than seniors. Perception of the AI agent averaged 4.12 with no significant difference between participant types.



Qualitative Results

🗣️ Support for Minority's Engagement
"It wasn't just me who had a different opinion." (P36)

⚖️ Empowerment & Balanced Dialogue
"AI gave a little more power to minority opinions." (P28)

🤔 Encouraging Critical Thinking
"AI made me think about options that had been overlooked." (P28)

⏳ Role Mismatch as Decision Phase Change
"It was good in the sense that it was kind of like a trigger for me... but the further it went on, the more I felt like I kind of tended to ignore it." (P15)

🤯 Better Decision Process VS Cognitive Load
"If the outcome is the same this way or that, then I think it's better to just make decisions without AI because it's better to use less energy." (P48)

Potential Design Implications

1 Timing of Interventions
Developing mechanisms for CAs to gauge real-time group dynamics, ensuring timely and context-aware inputs.

2 Clarity and Specificity
Leveraging retrieval-augmented generation (RAG) to provide specific, well-substantiated challenges that stimulate deeper critical thinking.

3 Group Reflection
Prompting personal insights and summarizing key discussion points to encourage diverse perspectives.

4 Argumentation Styles
Balancing assertiveness and inclusivity based on cohesion and diversity.

5 Dynamic Role Adaptation
Shifting between facilitator, supporter, and analyst roles to match group needs throughout discussions.

6 Balancing Critical Thinking and Group Dynamics
Monitoring engagement signals and adjusting intervention intensity to prevent cognitive overload.

DESIGN DEPT. | UNIST

Contact soohwanlee@unist.ac.kr

Acknowledgment

The authors gratefully acknowledge Dr. Angel Hsing-Chi Hwang and Dr. Oh-Sang Kwon for their assistance with experimental design and data analysis. 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). This work was also partially supported by the Technology Innovation Program (20015056, Commercialization design and development of Intelligent Product-Service System for personalized full silver life cycle care) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea).