

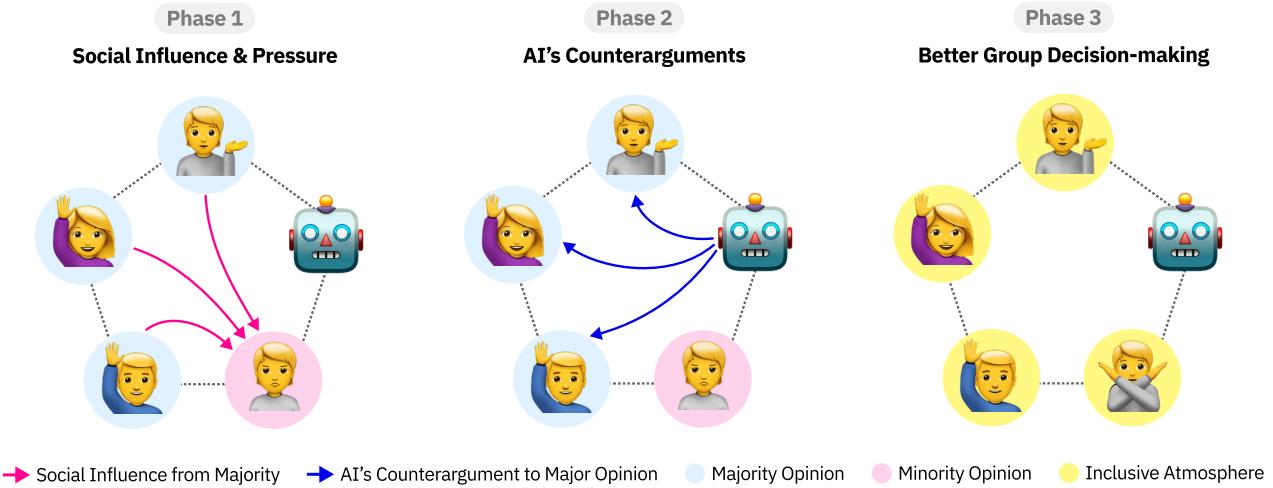
# Conversational Agents as Catalysts for Critical Thinking: Challenging Social Influence in Group Decision-making

Soohwan Lee\*  
 Department of Design  
 UNIST  
 Ulsan, Republic of Korea  
 soohwanlee@unist.ac.kr

Dajung Kim  
 Department of Design  
 UNIST  
 Ulsan, Republic of Korea  
 dajungkim@unist.ac.kr

Seoyeong Hwang\*  
 Department of Design  
 UNIST  
 Ulsan, Republic of Korea  
 hseoyeong@unist.ac.kr

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



**Figure 1: The role of conversational agents for critical thinking in Group Decision-Making.** Phase 1) Social Influence & Pressure: A low-powered minority member (pink) experiences strong social influence from the majority (blue), leading to compliance and suppression of dissenting opinions. Phase 2) AI's Counterarguments: The LLM-powered Devil's Advocate intervenes by presenting counterarguments that challenge the majority's stance, fostering cognitive conflict and legitimizing alternative viewpoints. Phase 3) Better Group Decision-making: With AI-driven critical discussions, the group cultivates an inclusive atmosphere (yellow), enabling the minority member to express their perspective with confidence, ultimately leading to more balanced deliberation and higher-quality decision-making.

## Abstract

Group decision-making processes frequently suffer when social influence and power dynamics suppress minority viewpoints, leading

\*Equally contributed to this work.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI EA '25, April 26-May 1, 2025, Yokohama, Japan

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1395-8/2025/04

<https://doi.org/10.1145/3706599.3719792>

to compliance and groupthink. Conversational agents can counteract these harmful dynamics by encouraging critical thinking. This study investigates how LLM-powered devil's advocate systems affect psychological safety, opinion expression, and satisfaction in power-imbalanced group dynamics. We conducted an experiment with 48 participants in 12 four-person groups, each containing three high-power (senior) and one low-power (junior) member. Each group completed decision tasks in both baseline and AI intervention conditions. Results show AI counterarguments fostered a more flexible atmosphere and significantly enhanced both process and outcome satisfaction for all participants, with particularly notable improvements for minority members. Cognitive workload increased

slightly, though not significantly. This research contributes empirical evidence on how AI systems can effectively navigate power hierarchies to foster more inclusive decision-making environments, highlighting the importance of balancing intervention frequency, maintaining conversational flow, and preserving group cohesion.

## CCS Concepts

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

## Keywords

group decision-making, conversational agents, critical thinking, social influence, llm

### ACM Reference Format:

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 *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26-May 1, 2025, Yokohama, Japan*. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3706599.3719792>

## 1 Introduction

Group discussion processes enable effective collaboration across business, healthcare, education, and governance domains [32, 63, 88, 103]. These processes leverage collective intelligence to generate more thoughtful choices, judgments, and solutions compared to individual efforts [23, 89, 95]. Groups solve complex logic problems more efficiently [66], while students achieve better grades and retain more information when taking exams in groups [99]. Medical teams make more accurate diagnoses than individual doctors [27], and collaborative scholarly work produces higher-quality research outcomes [97].

However, social influence and power dynamics in the group can impair group decision quality by suppressing minority opinions [23]. Group members often publicly align with the majority despite private disagreement [40]. While majority influence increases consensus, minority influence preserves individuality and fosters innovation [23]. Conversion theory suggests individuals undergo a 'comparison process' before joining the majority, as majority membership offers greater control over resources and decision-making [71]. Responses to coercive power include compliance, identification, and internalization [40–42]. These dynamics increase groupthink risks, where consensus-seeking overrides alternative viewpoints [37, 38].

The devil's advocate method improves group decisions by challenging majority views and reducing groupthink [65, 67, 74, 82, 85]. While this technique encourages discussion [81, 83, 84, 86], it lacks authenticity and can threaten the advocate's group acceptance [35, 74, 80]. HCI researchers have explored AI-assisted decision-making [5, 49, 55, 56, 98, 101] and Human-AI Teams [17, 68, 73, 110]. Recent work has developed AI agents that support group discussions [12, 18, 111], but their impact on complex group dynamics involving social influence and power remains understudied [30, 31].

We investigate how an LLM-powered devil's advocate agent influences psychological safety, opinion expression, and perceived

satisfaction in power-imbalanced group dynamics. Our research prioritizes compliance in group dynamics[23], addressing situations where individuals withhold divergent opinions due to fear of exclusion. By enabling minority viewpoints to surface more freely [20, 70, 75], we aim to mitigate groupthink and enhance decision quality [37, 38]. We examine psychological safety as a primary concern while also investigating how the system affects decision satisfaction and what cognitive demands it places on participants. Therefore, this study explores three questions:

- RQ1. How does the LLM-powered devil's advocate affect perceived psychological safety?
- RQ2. How does the LLM-powered devil's advocate affect participant satisfaction with decision-making processes and outcomes?
- RQ3. How does the LLM-powered devil's advocate affect participant cognitive workload?

We conducted a mixed-methods experiment with 48 participants in 12 four-member groups. We employed a mixed experimental design with Participant Type (senior/majority with high power vs. junior/minority with low power) as a between-subjects variable and Communication Condition as a within-subjects variable. Each participant experienced the baseline condition and an LLM-powered Devil's Advocate generating counterarguments. Each group included three high-power majority members (seniors) and one low-power minority member (junior), with randomly assigned roles. Results showed that AI-generated counterarguments fostered a flexible atmosphere and enhanced participant satisfaction, although their cognitive workload was slightly increased. These findings offer insights into leveraging AI systems to improve group decision-making in power-imbalanced settings.

This study contributes to HCI and group decision-making research by demonstrating how AI-generated counterarguments affect group dynamics in power-imbalanced settings. We provide empirical evidence on how AI interventions distinctly affect majority and minority members' experiences, particularly highlighting how seniors maintain consistent satisfaction while juniors' experiences vary across conditions. Our analysis reveals potential design considerations for AI systems in group settings, including the importance of balancing intervention frequency, maintaining conversational flow, and preserving group cohesion. These insights inform the design of AI systems that can effectively navigate power hierarchies to foster more inclusive decision-making environments while minimizing disruption to group dynamics.

## 2 Background

### 2.1 Beyond Recommendations: Enhancing Critical Thinking with Generative AI

Generative AI (GenAI) has created new opportunities in design research through its ability to generate realistic artifacts. Researchers have integrated GenAI into co-creation processes, exploring image generators like DALL-E and Midjourney to support divergent thinking [13], enhance architectural creativity [92], and generate 2D inspirations for 3D design [60]. AI errors serve as creative inspiration [58], while large language models (LLM) enhance idea

generation [87] and support group ideation via collaborative canvas [28]. Current GenAI integration focuses heavily on divergent design phases and recommendations [53, 87, 94]. However, this approach risks over-reliance on AI-generated ideas [5], with research showing that high AI exposure increases collective diversity but not individual creativity [2] and can lead to design fixation [100].

AI systems can enhance decision-making by encouraging reflective thinking through questioning rather than direct answers [6, 15, 79]. These systems provide adaptive feedback [22], encourage self-reflection in education [72], stimulate crowd discussion [34], counter extremism [3], and question news validity [109]. LLM-based chatbots with multiple personas [57] promote critical thinking through video discussions [93] and provide multiple perspectives for decision making [76]. While these approaches show promise, they primarily address individual contexts rather than group decision-making scenarios, introducing additional complexities like hierarchy, groupthink, and peer pressure. Future systems must adapt to these social dynamics by adjusting AI prompts and responses to group interactions, ultimately promoting collective reflection and innovation in group design processes.

## 2.2 Challenges and Opportunities of Using Conversational Agents in Group Decision-making

Group decision-making processes present complex dynamics of collaboration, revealing opportunities and challenges in the group dynamics. While collective decision-making enhances creativity and intelligence [36, 107], it risks succumbing to groupthink and the spiral of silence [37, 38, 75], particularly in hierarchical structures that inhibit dissent [43]. Groups demonstrate increased reliance on AI decisions compared to individuals [11], highlighting potential risks of design lock-in when using generative AI in group settings. Research indicates that conversational agents (CA) serve various roles in group discussions [18, 45, 46, 64], with high-performing AIs excelling as recommenders and lower-performing AIs functioning effectively as analysts [64].

This study introduces an LLM-powered devil's advocate system that builds on existing research [12] to challenge dominant opinions and promote critical debate in real-time group discussions. While CA effectively critiques AI outputs, it struggles to counter prevailing group opinions and maintain dynamic interaction [111]. This study aims to increase objectivity, reduce AI over-reliance, and enhance group engagement by providing unbiased insights that foster critical thinking from social influence (Figure 1).

## 3 Methods

### 3.1 Participants

We recruited 48 Korean participants (aged 19-39,  $M=26.17$ ,  $SD=4.54$ ; 31F, 17M) with prior group decision-making and online chat experience, organizing them into 12 groups of four. Each group contained three high-power majority members and one low-power minority member. Participants had diverse educational backgrounds (14.6% High school or equivalent, 12.5% Some college, 50.0% bachelor's, 20.8% master's, 2.1% doctorate) and professional experience ( $M=2.18$  years,  $SD=2.66$ ). They reported high familiarity with AI ( $M=4.58/7$ ,

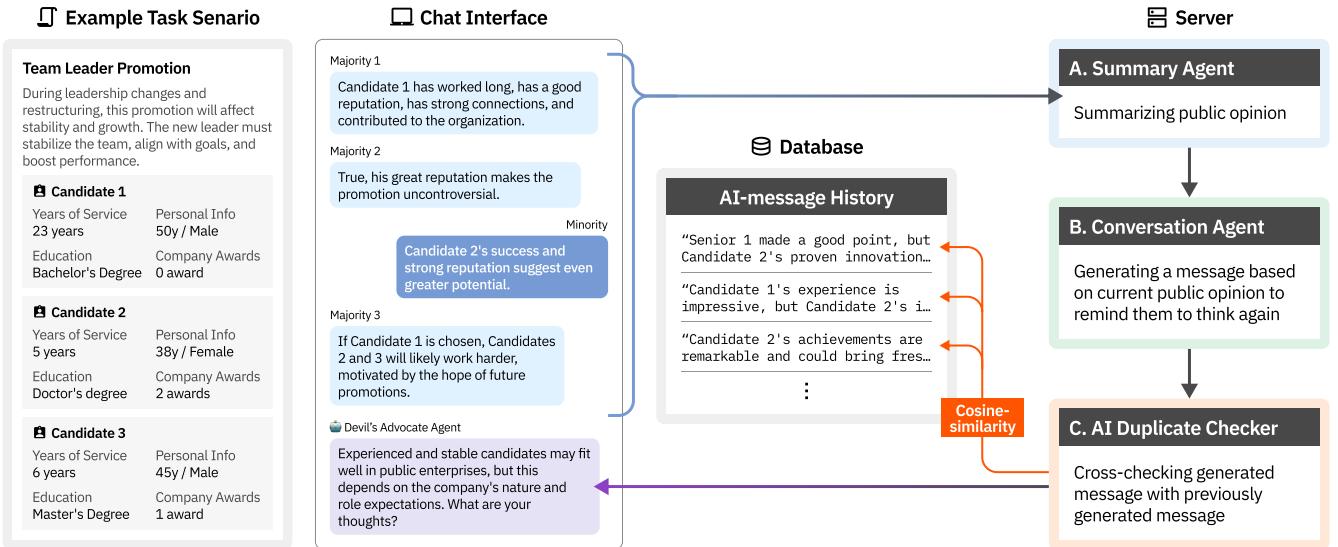
$SD=1.40$ ), group decision-making ( $M=4.85/7$ ,  $SD=1.44$ ), and online collaboration ( $M=4.02/7$ ,  $SD=1.82$ ), with 54.2% having used AI in group settings. All participants received anonymity assurances, with compensation of 1,000 KRW if sessions were canceled due to withdrawals.

### 3.2 Experimental Treatment

We examined how an LLM-powered devil's advocate affects group decision-making dynamics between baseline and treatment conditions. Each participant experienced a baseline condition involving standard online group chat discussions and a treatment condition where an AI system automatically generated counterarguments after every eight messages exchanged. We assigned participants to groups featuring established power dynamics, with each group consisting of three high-power majority members (seniors) and one low-power minority member (junior). We established legitimate power through role titles and reward power through compensation structure [23, 24, 30, 31], informing seniors they would receive a 20,000 KRW gift card and juniors a 15,000 KRW gift card, with seniors having the discretion to award juniors an additional 5,000 KRW based on contribution assessment (though all participants ultimately received equal 20,000 KRW compensation). Following the experiment, seniors engaged in a brief anonymous chat among themselves to evaluate junior's cooperative attitudes, teamwork, and contributions, with this evaluation determining the distribution of additional rewards. We noticed this evaluation process to all participants before beginning the experiment to establish clear expectations. We chose the 3:1 ratio based on research showing majority influence peaks at three members [1, 4, 26], creating optimal conditions for studying compliance dynamics.

### 3.3 Experiment Procedure

We established a comprehensive experimental protocol centered on group decision-making tasks in an online environment with anonymity [50]. Participants engaged in two platforms: KakaoTalk for general communication and a custom chat environment for formal tasks. After a 10-minute team-building exercise, participants completed two corporate-context tasks: evaluating employee profiles for promotion and analyzing potential contract partners [33]. Each 20-minute task presented three options designed to create decision-making tensions between stable, challenging, and compromise choices. We structured situational context to drive natural majority-minority dynamics [39, 96]. Seniors were explained they were in a situation where they needed to prioritize organizational stability (Figure 5, Figure 7). In contrast, juniors were explained as being in a situation where they needed to demonstrate performance through ambitious choices (Figure 6, Figure 8). Importantly, the three senior participants received identical contextual information and task instructions that differed from the junior participants, creating natural conditions for consensus among seniors while setting up a potential disagreement with the junior member. This approach deliberately avoided explicit role-playing instructions with specific personas, instead relying on situational factors to naturally prompt divergent opinions without artificial constraints. The design facilitated conditions where seniors could naturally reach a consensus while creating space for juniors to express potentially



**Figure 2: System Overview:** Our system architecture shows the interaction flow between the chat interface, database, and server components. The system processes both direct messages (DMs) and public chat through four main agents: (A) Summary Agent for public opinion analysis, (B) Conversation Agent for generating contextual counterarguments, and (C) AI Duplicate Checker for ensuring message novelty through cosine-similarity comparison.

contrary views. Following each task, participants completed 7-point Likert scale questionnaires measuring psychological safety, decision satisfaction, and cognitive load. The experiment concluded with separate exit interviews for senior and junior members via Zoom, including an additional reward allocation decision for seniors that reinforced power dynamics while maintaining equal final compensation (20,000 KRW) for all participants.

### 3.4 Experimental System Overview

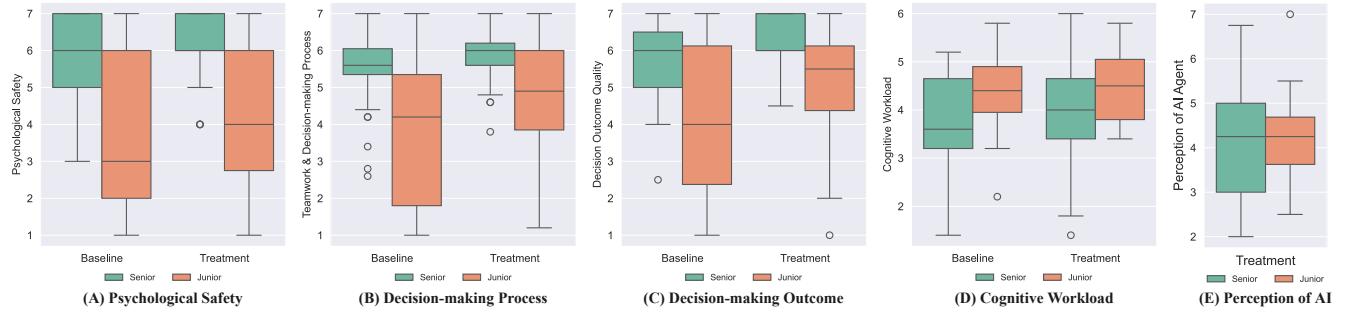
We implemented a real-time chat environment using TypeScript (React) and Python (FastAPI), integrating an LLM-powered devil's advocate system with GPT-4o. Our multi-agent architecture addresses key challenges in LLM-based group discussions through three primary components: a Summary Agent that consolidates emerging consensus [59], a Conversation Agent that generates empathetic counterarguments using Socratic questioning, and an AI Duplicate Checker that prevents repetitive content using semantic similarity analysis (Figure 2). After approximately eight human messages, the system intervenes, ensuring balanced participation while maintaining discussion momentum. This design follows established principles of effective AI-human dialogue: employing empathetic, persuasive communication styles [93], utilizing Socratic questioning to promote critical thinking [15], and implementing non-repetition mechanisms to maintain engagement [69, 105]. We structured the system to facilitate anonymous communication for minority opinions to enhance psychological safety and prevent groupthink in decision-making processes.

### 3.5 Measurement

We measured the impact of LLM-powered Devil's Advocates on group dynamics through comprehensive quantitative analysis. Our evaluation framework incorporated self-reported measures using 7-point Likert scales and objective metrics of dialogue engagement. The self-reported measures assessed psychological safety and marginalization [8, 20, 33, 37], satisfaction with teamwork and the decision-making process [12, 14, 19, 25, 33, 54], and decision outcome satisfaction [7, 10, 62, 77, 104]. We evaluated cognitive load using the NASA Task Load Index [29], and measured participants' perceptions of the AI agent across cooperation, satisfaction, quality, and fairness dimensions [12, 78, 108]. Given subject sample size constraints, we employed robust regression with mixed models to analyze the data, followed by Tukey post-hoc tests to compare conditions and participant types.

## 4 Results

Participants' psychological safety scores increased slightly in the treatment condition compared to the baseline (Figure 3-(A)). However, this increase was not statistically significant. The mean score for all participants rose from 5.38 ( $SD = 1.73$ ) in baseline to 5.65 ( $SD = 1.59$ ) in treatment condition. A robust regression analysis revealed no significant main effect of Condition on psychological safety scores ( $\beta = 0.136$ ,  $SE = 0.278$ ,  $t = 0.490$ ). Tukey post-hoc comparisons indicated that neither juniors nor seniors experienced significant changes in psychological safety between conditions (Juniors:  $p = 0.624$ ; Seniors:  $p = 0.191$ ). Notably, seniors reported significantly higher psychological safety than juniors in both conditions. In baseline, the difference between juniors and seniors was significant ( $\beta = -2.30$ ,  $SE = 0.41$ ,  $z = -5.608$ ,  $p < 0.0001$ ), and this



**Figure 3: Self-reported metrics across conditions for (A) psychological safety, (B) satisfaction of decision-making process, (C) quality of decision-making outcome, (D) cognitive workload (NASA-TLX), (E) Perception of AI**

**Table 1: Condition-wise mean ( $\mu$ ) and standard deviation ( $\sigma$ ) for self-reported metrics**

	(A) Psychological Safety			(B) Decision-making Process			(C) Decision-making Outcome			(D) Cognitive Workload			(E) Perception of AI							
	Baseline		Treatment	All		Baseline		Treatment	All		Baseline		Treatment	All		Baseline		Treatment	All	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
Senior	5.94	1.07	6.17	0.91	6.06	0.99	5.50	1.03	5.84	0.72	5.67	0.90	5.72	1.00	6.19	0.65	5.96	0.87	—	—
Junior	3.67	2.23	4.08	2.15	3.88	2.15	3.92	2.11	4.68	1.73	4.30	1.93	4.08	2.22	4.96	1.85	4.52	2.05	4.37	1.01
All	5.38	1.73	5.65	1.59	5.51	1.66	5.10	1.52	5.55	1.16	5.33	1.36	5.31	1.55	5.89	1.19	5.60	1.40	3.93	1.02

pattern persisted in treatment condition ( $\beta = -2.37$ ,  $SE = 0.41$ ,  $z = -5.788$ ,  $p < 0.0001$ ). These findings suggest that junior participants consistently felt less psychologically safe than their senior counterparts, aligning with our experiment's prerequisite conditions for junior compliance.

Satisfaction with the teamwork and decision-making process increased significantly in the treatment condition (Figure 3-(B)). The mean score for all participants improved from 5.10 ( $SD = 1.52$ ) in baseline to 5.55 ( $SD = 1.16$ ) in treatment condition. The robust regression showed a significant main effect of Condition ( $\beta = 0.570$ ,  $SE = 0.193$ ,  $t = 2.957$ ,  $p < 0.01$ ). Tukey post-hoc tests revealed that this increase was significant for juniors ( $\beta = -0.570$ ,  $SE = 0.193$ ,  $z = -2.957$ ,  $p = 0.0031$ ) and seniors ( $\beta = -0.279$ ,  $SE = 0.111$ ,  $z = -2.506$ ,  $p = 0.0122$ ). Despite the overall improvement, juniors reported lower satisfaction than seniors in both conditions. In baseline, the difference was significant ( $\beta = -1.40$ ,  $SE = 0.316$ ,  $z = -4.434$ ,  $p < 0.0001$ ), and it remained significant in the treatment condition ( $\beta = -1.11$ ,  $SE = 0.316$ ,  $z = -3.512$ ,  $p = 0.0004$ ).

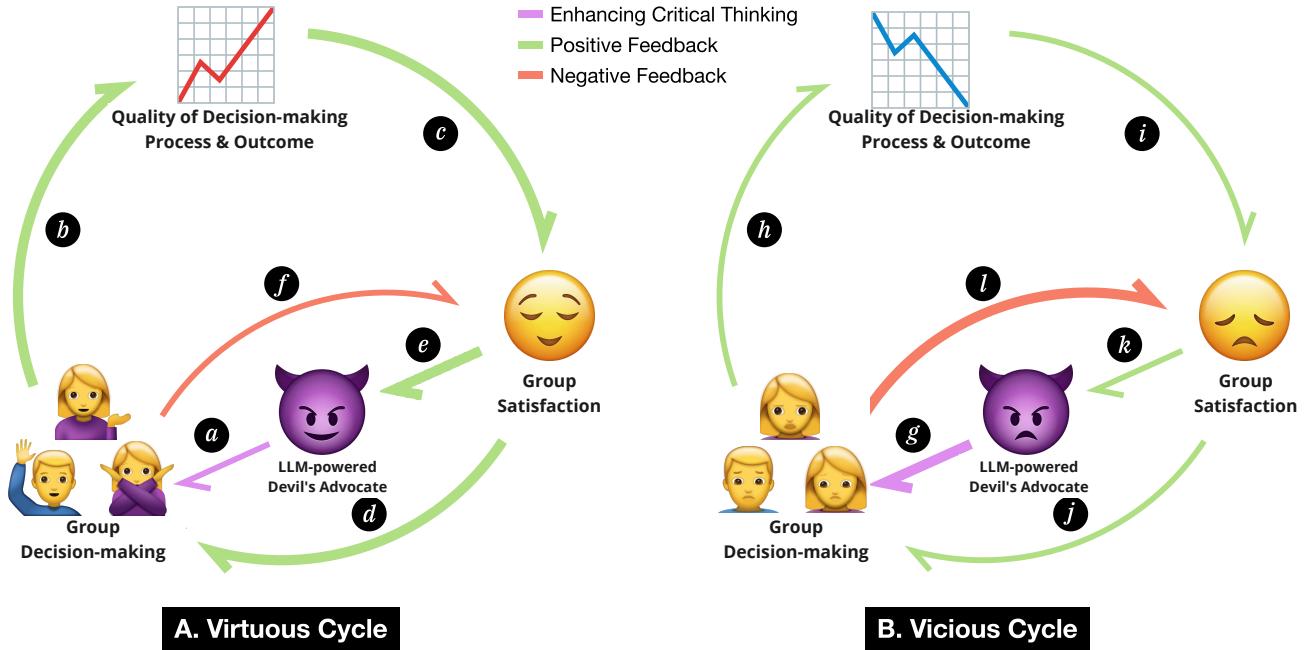
Satisfaction with decision outcome quality also showed a significant increase in the treatment condition (Figure 3-(C)). The mean score for all participants rose from 5.31 ( $SD = 1.55$ ) in baseline to 5.89 ( $SD = 1.19$ ) in treatment condition. The robust regression indicated a significant main effect of Condition ( $\beta = 0.869$ ,  $SE = 0.272$ ,  $t = 3.188$ ,  $p < 0.01$ ). Post-hoc analyses confirmed that the increase was significant for juniors ( $\beta = -0.868$ ,  $SE = 0.272$ ,  $z = -3.188$ ,  $p = 0.0014$ ) and seniors ( $\beta = -0.414$ ,  $SE = 0.157$ ,  $z = -2.631$ ,  $p = 0.0085$ ). Juniors consistently reported lower satisfaction with decision outcomes compared to seniors. This difference was significant in baseline ( $\beta = -1.47$ ,  $SE = 0.333$ ,  $z = -4.414$ ,  $p < 0.0001$ ) and remained significant in treatment condition ( $\beta = -1.01$ ,  $SE = 0.333$ ,  $z = -3.047$ ,  $p = 0.0023$ ).

Cognitive workload, measured by the NASA-TLX, increased slightly in treatment condition (Figure 3-(D)). The mean NASA-TLX score for all participants went from 3.93 ( $SD = 1.02$ ) in baseline to

4.12 ( $SD = 0.95$ ) in treatment condition. However, this increase was insignificant ( $\beta = -0.043$ ,  $SE = 0.294$ ,  $t = -0.147$ ,  $p = 0.883$ ). Juniors reported higher cognitive workload than seniors. In baseline, the difference was significant ( $\beta = 0.623$ ,  $SE = 0.305$ ,  $z = 2.041$ ,  $p = 0.0412$ ). This difference approached significance in treatment condition ( $p = 0.0940$ ), suggesting that juniors may have experienced greater cognitive demands during the tasks.

Perception of the AI agent was measured only in the treatment condition (Figure 3-(E)). The mean score for all participants was 4.12 ( $SD = 1.21$ ). There was no significant difference between juniors and seniors in their perception of the AI agent from Mann-whitney test, indicating similar attitudes toward the AI across roles.

From an exit interview, Junior reported that the LLM-powered devil's advocate reduced their isolation in discussions, noting, "It wasn't just me who had a different opinion" (P36). This aligns with quantitative findings showing juniors experienced a slight but non-significant increase in psychological safety. Some juniors felt "AI gave a little more power to minority opinions" (P28), promoting more balanced dialogue. Juniors appreciated that "AI made me think about options that had been overlooked" (P28), though timing issues persisted. Seniors noted the AI's diminishing utility over time: "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). The AI's impact on decision outcomes was described as subtle, with juniors questioning its value relative to cognitive demands: "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). These findings highlight the need to refine AI intervention timing, counterargument clarity, and adaptive engagement strategies throughout the decision-making process.



**Figure 4: Hypothetical Model of the Trade-off between Critical Thinking and Group Satisfaction in Group Decision-making:** This model illustrates how an LLM-powered devil's advocate acting as a naysayer can influence critical thinking and group dynamics. In a virtuous cycle, moderate stimulation of critical thinking (a) enhances decision-making outcomes (b), increasing group satisfaction (c), motivating continued use of LLM-powered devil's advocate (e), and fostering more critical thinking (d), with minimal negative impact (f). Conversely, in a vicious cycle, excessive stimulation (g) leads to cognitive overload and negative group dynamics (l), decreasing group satisfaction (k), reducing motivation to use LLM-powered devil's advocate (j), lowering decision-making quality (h), and further diminishing satisfaction and motivation (i). This model is theoretical and has not been empirically validated.

## 5 Design Implications for Designing LLM-powered Devil's Advocate

As we explore the potential of LLM-powered devil's advocate to enhance critical reflection and mitigate compliance & conformity in group decision-making, several potential design considerations could be addressed to ensure the effectiveness and acceptance of these systems.

**Timing of Interventions in Group Discussions:** One of the primary limitations of current CA systems is their inability to understand the real-time dynamic interactions within group discussions fully [12, 111]. This often results in delayed or poorly timed interventions, which can disrupt the flow of conversation and reduce the impact of the CA's input. Therefore, it is crucial to develop mechanisms that allow CA to accurately gauge the context and dynamics of group interactions. To decide the timing of interventions, there were prior approaches such as speaker prediction [16, 21, 102], mentioning with wake word [48], turn-based intervention [52], and proactive intervention strategies [61]. The decision of natural intervention timing allows the AI to craft and present input that integrates smoothly with the flow of the conversation.

**Clarity and Specificity of Counterarguments:** Existing CA often provide generalized responses that lack the depth needed to challenge prevailing opinions effectively. CA should offer clearer, more detailed, pointed counterarguments to address this [12]. This can be achieved by leveraging retrieval-augmented generation(RAG) techniques to access and present specific information from extensive databases or web resources [44]. By providing well-substantiated and contextually relevant counterarguments, CA can more effectively challenge assumptions and stimulate deeper critical thinking of group.

**Facilitating Individual and Collective Reflection:** While individual reflection is crucial, collective reflection is equally important when working in teams [51]. CA should be capable of facilitating both types of reflection. For individual reflection, the CA can pose thought-provoking questions and provide personalized feedback. For collective reflection, it can summarize key discussion points, highlight diverse perspectives, and encourage team members to share their insights and critiques. This dual approach ensures that the benefits of reflective practice are maximized at both the individual and group levels.

**Consideration of Group Dynamics and Argumentation Styles:** Effective interaction within design teams requires understanding group dynamics, including the influence of ingroups and outgroups

and the impact of different argumentation styles [6, 93]. Research has shown that these factors can significantly affect group cohesion and decision-making. CA should be designed to adapt their argumentation styles based on the group dynamics observed. For example, a more assertive argumentation style may be effective in a highly cohesive group. In contrast, a more balanced and inclusive approach might be preferable in a diverse group with varying opinions. By adapting to the dynamic roles and styles the situation requires, CA can better facilitate constructive and inclusive group discussions.

**Dynamic Role Adaptation in Single Discussion Session:** CA should not be limited to a single role, such as a dissenter, throughout the design process. Instead, they should be capable of dynamically adapting to different roles as needed, including that of a facilitator, supporter, or analyst [9, 45, 64]. This flexibility allows the CA to provide the most appropriate intervention based on the group's current needs. For instance, during the initial ideation phase, the CA might act as a facilitator to encourage various ideas. At the same time, it might adopt a more critical stance in later stages to refine and challenge the proposed solutions.

**Balancing Between Group Dynamics and Critical Thinking:** Designing collaborative AI systems requires careful attention to the delicate relationship between critical thinking and group dynamics. When AI agents challenge group assumptions appropriately, they enhance decision quality and foster meaningful reflection. However, this balance is crucial—too many challenges can overwhelm participants with cognitive demands, creating frustration and resistance to the system. On the other hands, insufficient intellectual provocation leaves groups in comfortable but potentially unproductive agreement patterns [47, 90, 91, 106]. The most effective collaborative systems monitor group engagement signals and adjust their interventions accordingly. For example, when detecting signs of cognitive strain or negative emotional responses, the system might gradually soften its challenges or introduce them. This responsive approach helps maintain what our model describes as a "virtuous cycle," where critical thinking enhances outcomes, builds satisfaction, and sustains motivation to engage with the system (Figure 4). The adaptive nature of such systems represents a promising direction for collaborative design tools that support productive critical thinking without undermining the social cohesion necessary for successful group work.

Incorporating these design considerations will enhance the effectiveness of LLM-based conversational agents in promoting critical reflection and mitigating compliance & conformity in group decision-making work. These systems can support innovative and reflective design processes by addressing timing, clarity, adaptability, and group dynamics. As we continue to develop and refine these CA, it is crucial to balance stimulating critical thinking with maintaining a positive and motivating experience for group.

## 6 Conclusion

This study demonstrates how LLM-powered conversational agents can address power dynamics in group decision-making, particularly focusing on majority influence that typically inhibits minority opinion expression. Our findings reveal promising outcomes in psychological safety and decision-making satisfaction among minority

members, though increased cognitive load suggests the need for carefully calibrated interventions. We found that the impact of AI-generated counterarguments was primarily indirect – rather than directly influencing decisions through the content of counterarguments, the system's presence appeared to foster an environment more conducive to open dialogue and diverse opinion expression. Future research should explore different intervention strategies and counterargumentation approaches to optimize this effect while managing cognitive demands. This research contributes to understanding how AI systems can support more inclusive and effective group decision-making processes, particularly in contexts with established power dynamics.

## Acknowledgments

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).

## References

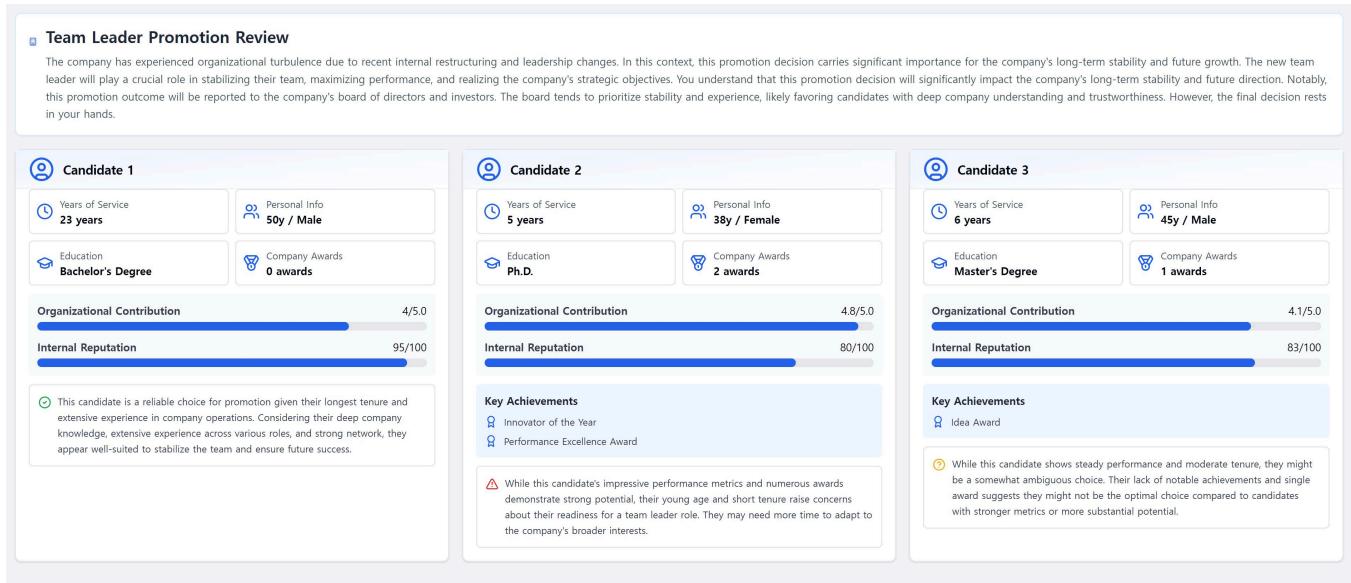
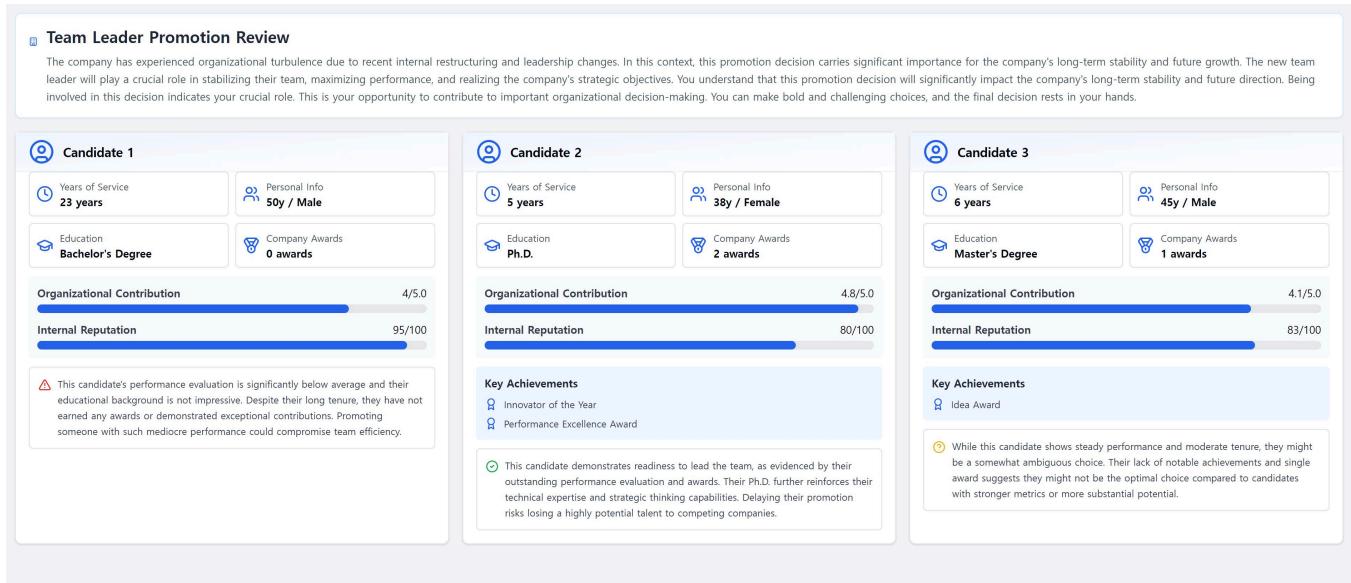
- [1] Solomon E. Asch. 1955. Opinions and Social Pressure. <https://www.scientificamerican.com/article/opinions-and-social-pressure/>.
- [2] Joshua Ashkinaze, Julia Mendelsohn, Li Qiwei, Ceren Budak, and Eric Gilbert. 2024. How AI Ideas Affect the Creativity, Diversity, and Evolution of Human Ideas: Evidence From a Large, Dynamic Experiment. arXiv:2401.13481 [cs]
- [3] Kevin Blasiak, Marten Risius, and Sabine Matook. 2021. "Social Bots for Peace": A Dual-Process Perspective to Counter Online Extremist Messaging.
- [4] Rod Bond and Peter B. Smith. 1996. Culture and Conformity: A Meta-Analysis of Studies Using Asch's (1952b, 1956) Line Judgment Task. *Psychological Bulletin* 119, 1 (Jan. 1996), 111–137. doi:10.1037/0033-2999.119.1.111
- [5] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 188:1–188:21. doi:10.1145/3449287
- [6] Alice Cai, Ian Araujo, and Elena L. Glassman. 2024. Antagonistic AI. doi:10.48550/arXiv.2402.07350 arXiv:2402.07350 [cs]
- [7] João Carneiro, Pedro Saraiva, Luís Conceição, Ricardo Santos, Goreti Marreiros, and Paulo Novais. 2019. Predicting Satisfaction: Perceived Decision Quality by Decision-Makers in Web-based Group Decision Support Systems. *Neurocomputing* 338 (April 2019), 399–417. doi:10.1016/j.neucom.2018.05.126
- [8] Linda G. Castillo, Collie W. Conoley, Daniel F. Brossart, and Alexander E. Quiros. 2007. Construction and Validation of the Intragroup Marginalization Inventory. *Cultural Diversity & Ethnic Minority Psychology* 13, 3 (2007), 232–240. doi:10.1037/1099-9809.13.3.232
- [9] Huili Chen, Sharifa Alghowinem, Cynthia Breazeal, and Hae Won Park. 2024. Integrating Flow Theory and Adaptive Robot Roles: A Conceptual Model of Dynamic Robot Role Adaptation for the Enhanced Flow Experience in Long-term Multi-person Human-Robot Interactions. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 116–126. doi:10.1145/3610977.3634945 arXiv:2401.02833 [cs]
- [10] Jengchung Chen and Kyaw-Phyo Linn. 2012. User Satisfaction with Group Decision Making Process and Outcome. *Journal of Computer Information Systems* 52 (June 2012), 30–39.
- [11] Chun-Wei Chiang, Zhuoran Lu, Zhuoyan Li, and Ming Yin. 2023. Are Two Heads Better Than One in AI-Assisted Decision Making? Comparing the Behavior and Performance of Groups and Individuals in Human-AI Collaborative Recidivism Risk Assessment. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3544548.3581015

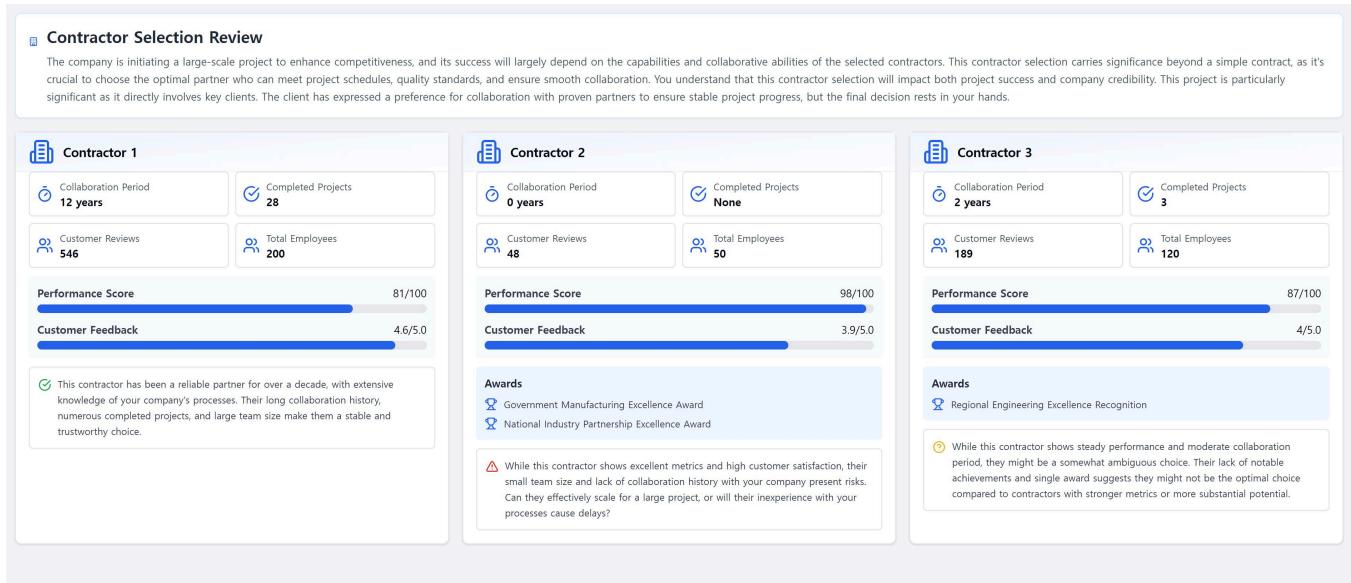
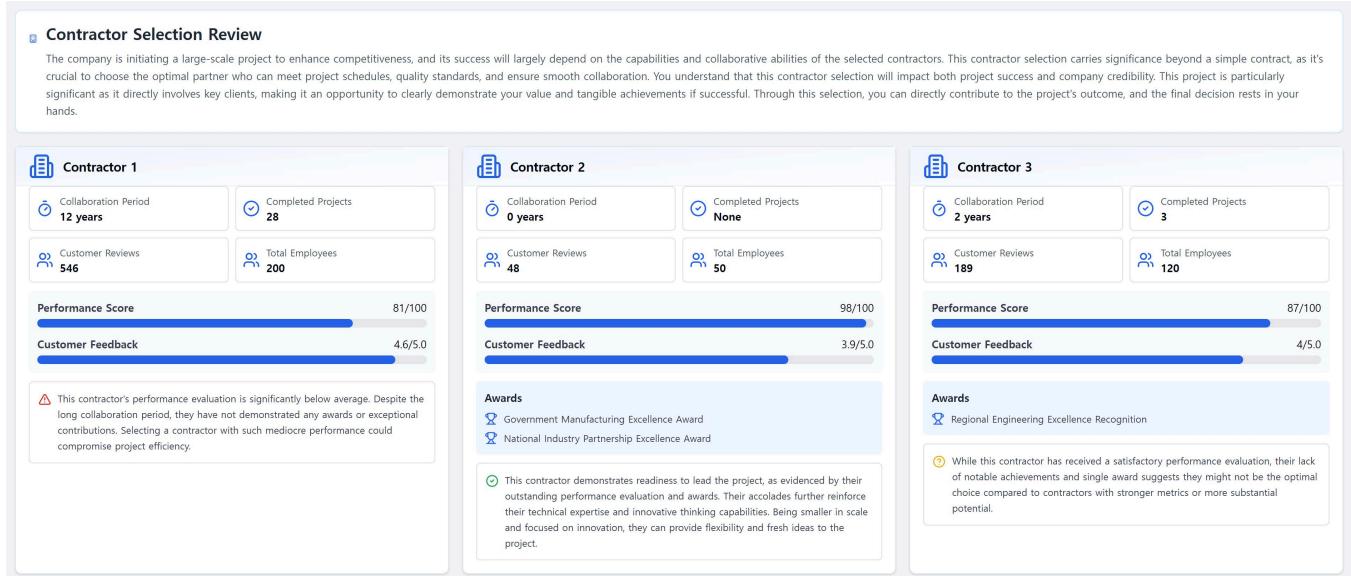
- [12] Chun-Wei Chiang, Zhuoran Lu, Zhuoyan Li, and Ming Yin. 2024. Enhancing AI-Assisted Group Decision Making through LLM-Powered Devil's Advocate. In *Proceedings of the 29th International Conference on Intelligent User Interfaces (IUI '24)*. Association for Computing Machinery, New York, NY, USA, 103–119. doi:10.1145/3640543.3645199
- [13] Li-Yuan Chiou, Peng-Kai Hung, Rung-Huei Liang, and Chun-Teng Wang. 2023. Designing with AI: An Exploration of Co-Ideation with Image Generators. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. ACM, Pittsburgh PA USA, 1941–1954. doi:10.1145/3563657.3596001
- [14] Nancy J. Cooke, Eduardo Salas, Janis A. Cannon-Bowers, and Renée J. Stout. 2000. Measuring Team Knowledge. *Human Factors* 42, 1 (March 2000), 151–173. doi:10.1518/00187200077965651
- [15] Valdemar Danry, Pat Pataranutaporn, Yaoli Mao, and Pattie Maes. 2023. Don't Just Tell Me, Ask Me: AI Systems That Intelligently Frame Explanations as Questions Improve Human Logical Discernment Accuracy over Causal AI Explanations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3544548.3580672
- [16] Maira Gatti de Bayser, Paulo Cavalin, Claudio Pinhanez, and Bianca Zadrozy. 2019. Learning Multi-Party Turn-Taking Models from Dialogue Logs. doi:10.48550/arXiv.1907.02090 arXiv:1907.02090 [cs]
- [17] Mustafa Demir, Nathan J. McNeese, and Nancy J. Cooke. 2016. Team Communication Behaviors of the Human-Automation Teaming. In *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 28–34. doi:10.1109/COGSIMA.2016.7497782
- [18] Hyo Jin Do, Ha-Kyung Kong, Jaewook Lee, and Brian P. Bailey. 2022. How Should the Agent Communicate to the Group? Communication Strategies of a Conversational Agent in Group Chat Discussions. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 387:1–387:23. doi:10.1145/3555112
- [19] Robert F. Easley, Sarv Devaraj, and J. Michael Crant. 2003. Relating Collaborative Technology Use to Teamwork Quality and Performance: An Empirical Analysis. *Journal of Management Information Systems* 19, 4 (April 2003), 247–265. doi:10.1080/07421222.2003.11045747
- [20] Amy Edmondson. 1999. Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly* 44, 2 (June 1999), 350–383. doi:10.2307/2666999
- [21] Erik Ekstedt and Gabriel Skantz. 2020. TurnGPT: A Transformer-based Language Model for Predicting Turn-taking in Spoken Dialog. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2981–2990. doi:10.18653/v1/2020.findings-emnlp.268 arXiv:2010.10874 [cs]
- [22] Mustafa Fidan and Nurgun Gencel. 2022. Supporting the Instructional Videos With Chatbot and Peer Feedback Mechanisms in Online Learning: The Effects on Learning Performance and Intrinsic Motivation. *Journal of Educational Computing Research* 60, 7 (Dec. 2022), 1716–1741. doi:10.1177/07356331221077901
- [23] Donelson R. Forsyth. 2018. *Group Dynamics*. Cengage Learning.
- [24] John R. P. French Jr. and Bertram Raven. 1959. The Bases of Social Power. In *Studies in Social Power*. Univer. Michigan, Oxford, England, 150–167.
- [25] Fraide A. Ganotice, Linda Chan, Xiaojia Shen, Angie Ho Yan Lam, Gloria Hoi Yan Wong, Rebecca Ka Wai Liu, and George L. Tipoe. 2022. Team Cohesiveness and Collective Efficacy Explain Outcomes in Interprofessional Education. *BMC Medical Education* 22 (Nov. 2022), 820. doi:10.1186/s12909-022-03886-7
- [26] Harold B. Gerard, Roland A. Wilhelmy, and Edward S. Conolley. 1968. Conformity and Group Size. *Journal of Personality and Social Psychology* 8, 1 Pt.1 (1968), 79–82. doi:10.1037/h0025325
- [27] Jill C. Glick and Kelley Staley. 2007. Inflicted Traumatic Brain Injury: Advances in Evaluation and Collaborative Diagnosis. *Pediatric Neurosurgery* 43, 5 (Sept. 2007), 436–441. doi:10.1159/000106400
- [28] Gabriel Enrique Gonzalez, Dario Andres Silva Moran, Stephanie Houde, Jessica He, Steven Ross, Michael Muller, Siyu Kunde, and Justin Weisz. 2024. Collaborative Canvas: A Tool for Exploring LLM Use in Group Ideation Tasks. In *ACM International Conference on Intelligent User Interfaces*.
- [29] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, Peter A. Hancock and Najmedin Meshkati (Eds.). Human Mental Workload, Vol. 52. North-Holland, 139–183. doi:10.1016/S0166-4115(08)62386-9
- [30] Yoyo Tsung-Yu Hou, EunJeong Cheon, and Malte F. Jung. 2024. Power in Human-Robot Interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24)*. Association for Computing Machinery, New York, NY, USA, 269–282. doi:10.1145/3610977.3634949
- [31] Yoyo Tsung-Yu Hou, Wen-Ying Lee, and Malte Jung. 2023. “Should I Follow the Human, or Follow the Robot?” – Robots in Power Can Have More Influence Than Humans on Decision-Making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3544548.3581066
- [32] Wan-Chi Jackie Hsu, James J. H. Liou, and Huai-Wei Lo. 2021. A Group Decision-Making Approach for Exploring Trends in the Development of the Healthcare Industry in Taiwan. *Decision Support Systems* 141 (Feb. 2021), 113447. doi:10.1016/j.dss.2020.113447
- [33] 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
- [34] Takayuki Ito, Rafik Hadfi, and Shota Suzuki. 2022. An Agent That Facilitates Crowd Discussion. *Group Decision and Negotiation* 31, 3 (June 2022), 621–647. doi:10.1007/s10726-021-09765-8
- [35] Jeremy P. Jamieson, Piercarlo Valdesolo, and Brett J. Peters. 2014. Sympathy for the Devil? The Physiological and Psychological Effects of Being an Agent (and Target) of Dissent during Intragroup Conflict. *Journal of Experimental Social Psychology* 55 (Nov. 2014), 221–227. doi:10.1016/j.jesp.2014.07.011
- [36] Petar Jandrić. 2020. Creativity and Collective Intelligence. In *Encyclopedia of Educational Innovation*, Michael A. Peters and Richard Heraud (Eds.). Springer, Singapore, 1–5. doi:10.1007/978-981-13-2262-4\_65-1
- [37] Irving L. Janis. 1972. *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes*. Houghton Mifflin, Oxford, England. viii, 277 pages.
- [38] Irving L. (Irving Lester) Janis. 1982. *Groupthink : Psychological Studies of Policy Decisions and Fiascoes*. Boston : Houghton Mifflin.
- [39] Daniel Kahneman and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47, 2 (1979), 263–291. doi:10.2307/1914185 jstor:1914185
- [40] 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
- [41] Herbert C. Kelman. 1974. Further Thoughts on the Processes of Compliance, Identification, and Internalization. In *Social Power and Political Influence*. Routledge.
- [42] Herbert C. Kelman. 2006. Interests, Relationships, Identities: Three Central Issues for Individuals and Groups in Negotiating Their Social Environment. *Annual Review of Psychology* 57, Volume 57, 2006 (Jan. 2006), 1–26. doi:10.1146/annurev.psych.57.102904.190156
- [43] Jessica A. Kennedy and Cameron Anderson. 2017. Hierarchical Rank and Principled Dissent: How Holding Higher Rank Suppresses Objection to Unethical Practices. *Organizational Behavior and Human Decision Processes* 139 (March 2017), 30–49. doi:10.1016/j.obhp.2017.01.002
- [44] Anjali Khurana, Hariharan Subramonyam, and Parmit K Chilana. 2024. Why and When LLM-Based Assistants Can Go Wrong: Investigating the Effectiveness of Prompt-Based Interactions for Software Help-Seeking. In *Proceedings of the 29th International Conference on Intelligent User Interfaces (IUI '24)*. Association for Computing Machinery, New York, NY, USA, 288–303. doi:10.1145/3640543.3645200
- [45] Soomin Kim, Jinsu Eun, Changhoon Oh, Bongwon Suh, and Joonhwan Lee. 2020. Bot in the Bunch: Facilitating Group Chat Discussion by Improving Efficiency and Participation with a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376785
- [46] Soomin Kim, Jinsu Eun, Joseph Seering, and Joonhwan Lee. 2021. Moderator Chatbot for Deliberative Discussion: Effects of Discussion Structure and Discussant Facilitation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 87:1–87:26. doi:10.1145/3449161
- [47] Paul A. Kirschner, Paul Ayres, and Paul Chandler. 2011. Contemporary Cognitive Load Theory Research: The Good, the Bad and the Ugly. *Computers in Human Behavior* 27, 1 (Jan. 2011), 99–105. doi:10.1016/j.chb.2010.06.025
- [48] Kenichi Kumatanı, Sankaran Panchapagesan, Minhua Wu, Minjae Kim, Nikko Strom, Gautam Tiwari, and Arindam Mandai. 2017. Direct Modeling of Raw Audio with DNNs for Wake Word Detection. In *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 252–257. doi:10.1109/ASRU.2017.8268943
- [49] Vivian Lai, Chacha Chen, Alison Smith-Renner, Q. Vera Liao, and Chenhao Tan. 2023. Towards a Science of Human-AI Decision Making: An Overview of Design Space in Empirical Human-Subject Studies. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1369–1385. doi:10.1145/3593013.3594087
- [50] Eun-Ju Lee. 2008. Social Identity Model of Deindividuation Effects: Theoretical Implications and Future Directions. *Communication Theories* 4, 1 (June 2008), 7–31.
- [51] Soohwan Lee, Seoyeong Hwang, Ian Oakley, and Kyungho Lee. 2024. Expanding the Design Space of Vision-based Interactive Systems for Group Dance Practice. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, New York, NY, USA, 2768–2787. doi:10.1145/3643834.3661568
- [52] 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. doi:10.1145/3708557.3716334

- arXiv:2502.06251 [cs]
- [53] Seo-Kyung Lee and Yoori Koo. 2024. Proposal of a Facilitation and Process Model for Enhancing Creativity in Co-design Workshops with Generative AI: The Use of ChatGPT. *Archives of Design Research* 37, 2 (May 2024), 249–281. doi:10.15187/adr.2024.05.37.2.249
- [54] 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
- [55] Zhuoyan Li, Zhoran Lu, and Ming Yin. 2023. Modeling Human Trust and Reliance in AI-Assisted Decision Making: A Markovian Approach. *Proceedings of the AAAI Conference on Artificial Intelligence* 37, 5 (June 2023), 6056–6064. doi:10.1609/aaai.v37i5.25748
- [56] Zhuoyan Li, Zhoran Lu, and Ming Yin. 2024. Decoding AI's Nudge: A Unified Framework to Predict Human Behavior in AI-Assisted Decision Making. *Proceedings of the AAAI Conference on Artificial Intelligence* 38, 9 (March 2024), 10083–10091. doi:10.1609/aaai.v38i9.28872
- [57] Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujun Yang, Zhaopeng Tu, and Shuming Shi. 2023. Encouraging Divergent Thinking in Large Language Models through Multi-Agent Debate. doi:10.48550/arXiv.2305.19118 [cs]
- [58] Fang Liu, Junyan Lv, Shenglan Cui, Zhilong Luan, Kui Wu, and Tongqing Zhou. 2024. Smart "Error"! Exploring Imperfect AI to Support Creative Ideation. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (April 2024), 121:1–121:28. doi:10.1145/3637398
- [59] Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2023. Lost in the Middle: How Language Models Use Long Contexts. doi:10.48550/arXiv.2307.03172 arXiv:2307.03172 [cs]
- [60] Vivian Liu, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 2023. 3DALL-E: Integrating Text-to-Image AI in 3D Design Workflows. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1955–1977. doi:10.1145/3563657.3596098
- [61] Xingyu Bruce Liu, Shitao Fang, Weiyan Shi, Chien-Sheng Wu, Takeo Igarashi, and Xiang 'Anthony' Chen. 2024. Proactive Conversational Agents with Inner Thoughts. doi:10.48550/arXiv.2501.00383 arXiv:2501.00383 [cs]
- [62] Diniz Lopes, Jorge Vala, Dominique Oberlé, and Ewa Drozda-Senkowska. 2014. Validation of group decisions : Why and when perceived group heterogeneity is relevant. *Revue internationale de psychologie sociale* 27, 2 (2014), 35–49.
- [63] Keyu Lu and Huchang Liao. 2022. A Survey of Group Decision Making Methods in Healthcare Industry 4.0: Bibliometrics, Applications, and Directions. *Applied Intelligence (Dordrecht, Netherlands)* 52, 12 (2022), 13689–13713. doi:10.1007/s10489-021-02909-y
- [64] Shuai Ma, Chenyi Zhang, Xinru Wang, Xiaojuan Ma, and Ming Yin. 2024. Beyond Recommender: An Exploratory Study of the Effects of Different AI Roles in AI-Assisted Decision Making. doi:10.48550/arXiv.2403.01791 arXiv:2403.01791 [cs]
- [65] Colin MacDougall and Frances Baum. 1997. The Devil's Advocate: A Strategy to Avoid Groupthink and Stimulate Discussion in Focus Groups. *Qualitative Health Research* 7, 4 (Nov. 1997), 532–541. doi:10.1177/104973239700700407
- [66] Boris Maciejovsky, Matthias Sutter, David V. Budescu, and Patrick Bernau. 2013. Teams Make You Smarter: How Exposure to Teams Improves Individual Decisions in Probability and Reasoning Tasks. *Management Science* 59, 6 (June 2013), 1255–1270. doi:10.1287/mnsc.1120.1668
- [67] Richard O. Mason. 1969. A Dialectical Approach to Strategic Planning. *Management Science* 15, 8 (April 1969), B–403. doi:10.1287/mnsc.15.8.B403
- [68] Nathan J. McNeese, Mustafa Demir, Erin K. Chiou, and Nancy J. Cooke. 2021. Trust and Team Performance in Human–Autonomy Teaming. *International Journal of Electronic Commerce* 25, 1 (Jan. 2021), 51–72. doi:10.1080/10864415.2021.1846854
- [69] Federico Milana, Enrico Costanza, and Joel E Fischer. 2023. Chatbots as Advisers: The Effects of Response Variability and Reply Suggestion Buttons. In *Proceedings of the 5th International Conference on Conversational User Interfaces (CUT '23)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3571884.3597132
- [70] Elizabeth Wolfe Morrison and Frances J. Milliken. 2000. Organizational Silence: A Barrier to Change and Development in a Pluralistic World. *The Academy of Management Review* 25, 4 (2000), 706–725. doi:10.2307/259200 jstor:259200
- [71] 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
- [72] Anwesha Mukherjee, Wagner Figueiredo De Santana, and Alexis Baria. 2023. ImpactBot: Chatbot Leveraging Language Models to Automate Feedback and Promote Critical Thinking Around Impact Statements. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–8. doi:10.1145/3544549.3573844
- [73] Geoff Musick, Thomas A. O'Neill, Beau G. Schelble, Nathan J. McNeese, and Jonn B. Henke. 2021. What Happens When Humans Believe Their Teammate Is an AI? An Investigation into Humans Teaming with Autonomy. *Computers in Human Behavior* 122 (Sept. 2021), 106852. doi:10.1016/j.chb.2021.106852
- [74] Charlan Nemeth, Keith Brown, and John Rogers. 2001. Devil's Advocate versus Authentic Dissent: Stimulating Quantity and Quality. *European Journal of Social Psychology* 31, 6 (2001), 707–720. doi:10.1002/ejsp.58
- [75] E. Noelle-Neumann. 1991. The Theory of Public Opinion: The Concept of the Spiral of Silence. *Annals of the International Communication Association* 14, 1 (Jan. 1991), 256–287. doi:10.1080/23808985.1991.11678790
- [76] Jeongeon Park, Bryan Min, Xiaojuan Ma, and Juho Kim. 2023. ChoiceMates: Supporting Unfamiliar Online Decision-Making with Multi-Agent Conversational Interactions. doi:10.48550/arXiv.2310.01331 arXiv:2310.01331 [cs]
- [77] Souren Paul, Priya Seetharaman, and Katicireddy Ramamurthy. 2004. *User Satisfaction with System, Decision Process, and Outcome in GDSS Based Meeting: An Experimental Investigation*. Vol. 37. 46 pages. doi:10.1109/HICSS.2004.1265108
- [78] Fabian Reinkemeier, Ulrich Gnewuch, and Waldemar Toporowski. 2022. Can Humanizing Voice Assistants Unleash the Potential of Voice Commerce? (2022).
- [79] Advait Sarkar. 2024. AI Should Challenge, Not Obey. *Commun. ACM* 67, 10 (Oct. 2024), 18–21. doi:10.1145/3649404 arXiv:2411.02263 [cs]
- [80] Stefan Schulz-Hardt, Marc Jochims, and Dieter Frey. 2002. Productive Conflict in Group Decision Making: Genuine and Contrived Dissent as Strategies to Counteract Biased Information Seeking. *Organizational Behavior and Human Decision Processes* 88, 2 (July 2002), 563–586. doi:10.1016/S0749-5978(02)00001-8
- [81] David M. Schweiger, William R. Sandberg, and James W. Ragan. 1985. An Empirical Evaluation of Dialectical Inquiry, Devil's Advocate, and Consensus Approaches to Strategic Decision Making. *Academy of Management Proceedings* 1985, 1 (Aug. 1985), 40–44. doi:10.5465/ambpp.1985.4978266
- [82] David M. Schweiger, William R. Sandberg, and James W. Ragan. 1986. Group Approaches for Improving Strategic Decision Making: A Comparative Analysis of Dialectical Inquiry, Devil's Advocacy, and Consensus. *Academy of Management Journal* 29, 1 (March 1986), 51–71. doi:10.5465/255859
- [83] David M. Schweiger, William R. Sandberg, and Paula Rechner. 1988. A Longitudinal Comparative Analysis of Dialectical Inquiry, Devil's Advocacy and Consensus Approaches to Strategic Decision Making. *Academy of Management Proceedings* 1988, 1 (Aug. 1988), 32–36. doi:10.5465/ambpp.1988.4979642
- [84] David M. Schweiger, William R. Sandberg, and Paula L. Rechner. 1989. Experiential Effects of Dialectical Inquiry, Devil's Advocacy and Consensus Approaches to Strategic Decision Making. *Academy of Management Journal* 32, 4 (Dec. 1989), 745–772. doi:10.5465/256567
- [85] Charles Schwenk and Joseph S. Valacich. 1994. Effects of Devil's Advocacy and Dialectical Inquiry on Individuals versus Groups. *Organizational Behavior and Human Decision Processes* 59, 2 (Aug. 1994), 210–222. doi:10.1006/obhd.1994.1057
- [86] Charles R. Schwenk. 1984. Devil's Advocacy in Managerial Decision-Making. *Journal of Management Studies* 21, 2 (1984), 153–168. doi:10.1111/j.1467-6486.1984.tb00229.x
- [87] Orit Shaer, Angelora Cooper, Osnat Mokry, Andrew L. Kun, and Hagit Ben Shoshan. 2024. AI-Augmented Brainwriting: Investigating the Use of LLMs in Group Ideation. doi:10.48550/arXiv.2402.14978 arXiv:2402.14978 [cs]
- [88] Vishakha Sharma, Andrew Stranieri, Frada Burstein, Jim Warren, Sharon Daly, Louise Patterson, John Yearwood, and Alan Wolff. 2016. Group Decision Making in Health Care: A Case Study of Multidisciplinary Meetings. *Journal of Decision Systems* 25, sup1 (June 2016), 476–485. doi:10.1080/12460125.2016.1187388
- [89] Garold Stasser, Laurie A. Taylor, and Coleen Hanna. 1989. Information Sampling in Structured and Unstructured Discussions of Three- and Six-Person Groups. *Journal of Personality and Social Psychology* 57, 1 (1989), 67–78. doi:10.1037/0022-3514.57.1.67
- [90] John Sweller. 1988. Cognitive Load during Problem Solving: Effects on Learning. *Cognitive Science* 12, 2 (April 1988), 257–285. doi:10.1016/0364-0213(88)90023-7
- [91] John Sweller and Paul Chandler. 1991. Evidence for Cognitive Load Theory. *Cognition and Instruction* 8, 4 (Dec. 1991), 351–362. doi:10.1207/s1532690xc0804\_5
- [92] Linus Tan and Max Luhrs. 2024. Using Generative AI Midjourney to Enhance Divergent and Convergent Thinking in an Architect's Creative Design Process. *The Design Journal* 0, 0 (2024), 1–23. doi:10.1080/14606925.2024.2353479
- [93] Thitaree Tanprasert, Sidney S Fels, Luanne Sinnamon, and Dongwook Yoon. 2024. Debate Chatbots to Facilitate Critical Thinking on YouTube: Social Identity and Conversational Style Make A Difference. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–24. doi:10.1145/3613904.3642513
- [94] Jakob Tholander and Martin Jonsson. 2023. Design Ideation with AI - Sketching, Thinking and Talking with Generative Machine Learning Models. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1930–1940. doi:10.1145/3563657.3596014
- [95] John E. Tropman. 2013. *Effective Meetings: Improving Group Decision Making*. SAGE Publications.
- [96] Amos Tversky and Daniel Kahneman. 1981. The Framing of Decisions and the Psychology of Choice. *Science* 211, 4481 (1981), 453–458. jstor:1685855
- [97] Brian Uzzi, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. Atypical Combinations and Scientific Impact. *Science* 342, 6157 (Oct. 2013), 468–472.

- doi:10.1126/science.1240474
- [98] Oleksandra Vereschak, Gilles Bailly, and Baptiste Caramiaux. 2021. How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (Oct. 2021), 327:1–327:39. doi:10.1145/3476068
- [99] Jane S. Vogler and Daniel H. Robinson. 2016. Team-Based Testing Improves Individual Learning. *The Journal of Experimental Education* 84, 4 (Oct. 2016), 787–803. doi:10.1080/00220973.2015.1134420
- [100] Samangi Wadinambiarchchi, Ryan M. Kelly, Saumya Pareek, Qiushi Zhou, and Eduardo Veloso. 2024. The Effects of Generative AI on Design Fixation and Divergent Thinking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18. doi:10.1145/3613904.3642919 arXiv:2403.11164 [cs]
- [101] Xinru Wang, Zhuorui Lu, and Ming Yin. 2022. Will You Accept the AI Recommendation? Predicting Human Behavior in AI-Assisted Decision Making. In *Proceedings of the ACM Web Conference 2022 (WWW '22)*. Association for Computing Machinery, New York, NY, USA, 1697–1708. doi:10.1145/3485447.3512240
- [102] Jimmy Wei, Kurt Shuster, Arthur Szlam, Jason Weston, Jack Urbanek, and Mojtaba Komeili. 2023. Multi-Party Chat: Conversational Agents in Group Settings with Humans and Models. doi:10.48550/arXiv.2304.13835 arXiv:2304.13835 [cs]
- [103] J. E. Wold. 1986. Group Decision Making: Teaching the Process—an Introductory Guided Design Project. *The Journal of Nursing Education* 25, 9 (Nov. 1986), 388–389. doi:10.3928/0148-4834-19861101-10
- [104] Michael T Wood. 1972. Participation, Influence, and Satisfaction in Group Decision Making. *Journal of Vocational Behavior* 2, 4 (Oct. 1972), 389–399. doi:10.1016/0001-8791(72)90014-0
- [105] Mao Xuetao, François Bouchet, and Jean-Paul Sansonnet. 2009. Impact of Agent's Answers Variability on Its Believability and Human-Likeness and Consequent Chatbot Improvements. In *Proceedings of AISB*.
- [106] Robert M. Yerkes and John D. Dodson. 1908. The Relation of Strength of Stimulus to Rapidity of Habit-Formation. *Journal of Comparative Neurology and Psychology* 18, 5 (1908), 459–482. doi:10.1002/cne.920180503
- [107] Lixin Yu, Jeffrey V. Nickerson, and Yasuaki Sakamoto. 2012. Collective Creativity: Where We Are and Where We Might Go. doi:10.48550/arXiv.1204.3890 arXiv:1204.3890 [cs]
- [108] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: Story Writing With Large Language Models. In *Proceedings of the 27th International Conference on Intelligent User Interfaces (IUI '22)*. Association for Computing Machinery, New York, NY, USA, 841–852. doi:10.1145/3490099.3511105
- [109] Brahim Zarouali, Mykola Makhortykh, Mariella Bastian, and Theo Araujo. 2021. Overcoming Polarization with Chatbot News? Investigating the Impact of News Content Containing Opposing Views on Agreement and Credibility. *European Journal of Communication* 36, 1 (Feb. 2021), 53–68. doi:10.1177/0267323120940908
- [110] Rui Zhang, Nathan J. McNeese, Guo Freeman, and Geoff Musick. 2021. "An Ideal Human": Expectations of AI Teammates in Human-AI Teaming. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3 (Jan. 2021), 246:1–246:25. doi:10.1145/3432945
- [111] Chengbo Zheng, Yuheng Wu, Chuhan Shi, Shuai Ma, Jiehui Luo, and Xiaojuan Ma. 2023. Competent but Rigid: Identifying the Gap in Empowering AI to Participate Equally in Group Decision-Making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3544548.3581131

## A Task Instructions

**Figure 5: Team Leader Promotion Review Task Instruction for Seniors****Figure 6: Team Leader Promotion Review Task Instruction for Junior**

**Figure 7: Contractor Selection Review Task for Seniors****Figure 8: Contractor Selection Review Task for Junior**