Investigating LLM-Powered Minority Support in Power-Imbalanced Group Decision-Making: Counterargument and Mediation as Intervention Strategies

ANONYMOUS AUTHOR(S)

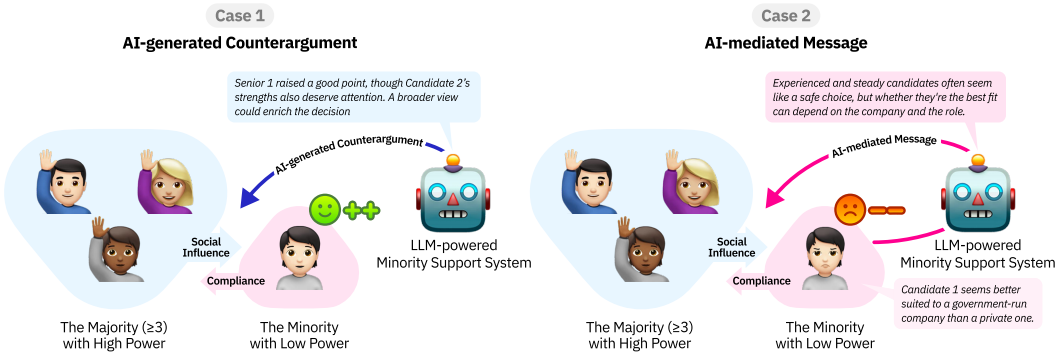


Fig. 1. LLM-powered minority support system mediates between majority and minority group members through two designs. The AI-generated Counterargument (AIGC) condition generated counterpoints to the majority consensus, broadening the discussion and improving the group atmosphere. The AI-mediated Message (AIMM) condition paraphrased minority input for anonymity but often reduced participants' psychological safety and satisfaction.

Minority opinions are often suppressed in power-imbalanced group decision-making due to social pressure to comply with the majority. To address this problem, we developed an LLM-powered minority support system that aimed to foster attention to minority views through either AI-generated counterarguments or AI-mediated messages. We conducted a mixed-method experiment with 96 participants in 24 groups, comparing minority members' experiences across baseline, AI-counterargument, and AI-mediated message conditions. Our findings revealed a nuanced trade-off: AI-generated counterarguments fostered a more flexible atmosphere and enhanced satisfaction, while AI-mediated messaging increased minority participation but unexpectedly reduced their psychological safety. This research contributes empirical evidence on how different AI implementations affect group dynamics, identifies a critical support paradox between participation and psychological safety, provides design implications for future systems, and highlights ethical challenges in implementing AI-mediated communication in hierarchical settings. These insights advance understanding of designing more equitable AI support for power-imbalanced group decision-making.

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

Additional Key Words and Phrases: group decision-making, conversational agents, critical thinking, social influence, llm

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

ACM Reference Format:
Anonymous Author(s). 2018. Investigating LLM-Powered Minority Support in Power-Imbalanced Group Decision-Making: Counterargument and Mediation as Intervention Strategies. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 35 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Power imbalances in group decision-making frequently suppress minority perspectives, restrict the diversity of ideas, and weaken overall outcomes [49]. Compliance pressures often lead individuals to publicly align with the majority despite private disagreement, undermining psychological safety and discouraging meaningful participation from less-empowered members [53, 69]. While collaborative processes help groups solve complex challenges, support individual learning, and lead to more accurate and creative outcomes across domains such as healthcare, education, and research [35, 64, 86, 87], these benefits could be worsened and diminished, especially when hierarchical structures or conformity suppress dissent. In an effort to address barriers to have more balanced, active participation, recent studies have introduced AI-powered interventions that facilitate idea generation, consensus-building, and alternative viewpoint introduction to encourage engagement and mitigate groupthink [13, 54, 81]. These interventions typically operate through four underlying mechanisms: by regulating atmosphere, using affective cues to foster supportive group atmosphere; by balancing participation, prompting quieter members and curbing dominance to promote equity; and by diversifying perspectives, ensuring that groups consider a broader set of alternatives.

However, most AI-powered interventions are designed around the assumption that all members have equal standing and fair opportunities to contribute to participation and consensus-building. The existing approaches may improve aggregate engagement, but it is underexplored how they effectively address the challenge of supporting and amplifying minority voices under conditions of unequal power. Research on supporting and amplifying minority voices has also explored the potential of systematic interventions, such as anonymous feedback platforms and automated counterargument generators, to address these issues [22, 46, 57]. Yet, there is a significant gap, especially in how to represent authentic minority voices or adequately safeguard psychological safety and anonymity, which is crucial for minorities' experiences in group decision-making.

To address this gap, this research empirically investigates how LLM-powered minority support interventions can foster psychologically safe and equitable environments for minority members expressing dissent in hierarchical decision-making settings. We designed two intervention conditions: an AI-generated Counterargument (AIGC) condition, where the AI produces alternative viewpoints automatically for every four turns and counterpoints to stimulate discussion, and an AI-mediated Message (AIMM) condition, which extends AIGC by allowing minority members to send private inputs to the AI that are then paraphrased and introduced as system contributions. This dual functionality allows minority members to express dissenting views without fear of social consequences, potentially enabling individuals experiencing compliance pressure to feel more psychologically secure, participate more actively, and gain greater satisfaction with group decision processes. We examine the influence of these conditions on group decision making under compliance pressure with the four factors. To what extent do these conditions affect **Perceived psychological safety (RQ1)**, **Engagement (RQ2)**, **Satisfaction with decision-making processes and outcomes (RQ3)**, and **Cognitive workload of participants in hierarchical discussions (RQ4)**?

To investigate these questions, we conducted a mixed-methods study with 96 participants organized into 24 groups of four members each. Each group included three members assigned as high-power majority (seniors) and one member as low-power minority (junior), with roles randomly assigned. We employed a mixed experimental design with participant type (majority vs.

minority) as a between-subjects variable and type of LLM-powered minority support (AIGC, AIMM) as a partially within-subjects variable. Our findings indicate that AI-generated counterarguments improved group atmosphere and participant satisfaction while enabling the minority member to submit anonymous opinions through AI increased discussion but reduced psychological safety and satisfaction for minority participants. These results underscore the complex trade-offs between amplifying minority voices and maintaining psychological safety when using LLM-powered support for minority in group decision-making.

This research makes four key contributions to CSCW. First, we empirically demonstrate how different LLM-powered minority support interventions affect minorities' experiences under power-imbalanced group decision-making, revealing that AI-generated counterarguments foster better discussions and satisfaction compared to AI-mediated minority messaging. Second, we identify a critical trade-off in minority support: while AI-mediated messaging increases minority participation, it simultaneously decreases their psychological safety, challenging assumptions about anonymous communication channels. Third, we provide design implications for future AI-powered minority support, focusing on agency, anonymity, and transparency. Finally, we highlight ethical considerations for implementing AI systems in hierarchical decision-making contexts. The remainder of this paper is structured as follows: we review related work on group decision-making and AI-mediated communication, describe methodology, present findings, discuss implications, and conclude with limitations and future research directions.

2 Related Work

2.1 The Impact of Social Influence and Power on Group Decision-making

Group decision-making leverages collective intelligence to produce superior outcomes across various domains [35, 64, 86], but these processes are significantly shaped by social influence and power dynamics [53, 69]. Social influence theory suggests that individuals tend to adjust their behavior to meet social demands, with the majority's opinions exerting particularly strong pressure on those with less power in the group. Moscovici's conversion theory specifically explains that multiple influences trigger a comparison process resulting in compliance - a form of conformity where individuals outwardly agree while maintaining private disagreement [69]. This compliance is typically direct, immediate, and temporary, serving as a coping mechanism in power-imbalanced situations rather than reflecting genuine belief change.

Power dynamics become especially problematic in hierarchical settings where power imbalances are formalized through reward and legitimate power structures [28]. Kelman's framework provides particular insight here, identifying compliance as an initial response to power where individuals conform primarily to avoid repercussions or gain rewards, rather than from genuine conviction [53]. This dynamic is especially evident among minority members, who are often treated as outgroup members and experience isolation. The effect is particularly pronounced when the size disparity between majority and minority groups is substantial. The resulting self-censorship triggers a cascade of negative effects: as minority voices are silenced, groups lose access to diverse perspectives that could enhance decision quality, ultimately leading to groupthink, where the desire for consensus overrides critical evaluation of alternatives [48, 49, 52].

Traditional social psychology offers several approaches to address compliance and prevent its progression to groupthink in group settings. The devil's advocate technique addresses groupthink by assigning someone to challenge prevailing viewpoints, stimulating opinion diversity [63, 65, 71, 78, 79]. However, this method doesn't directly solve the underlying compliance issues and faces significant limitations: designated advocates often lack authentic dissenting perspectives,

risk social ostracism when challenging powerful members, and cannot accurately represent unexpressed minority viewpoints [47, 71, 77]. Other compliance-reduction strategies include leadership interventions to create psychologically safe environments [24] and anonymous feedback channels [50]. However, these methods face practical challenges: leadership interventions rely on the leader's skills, and anonymous feedback often fails in small groups where unique opinions reveal identities. This research addresses these limitations by exploring how conversational AI might provide both psychological safety and true anonymity while preserving the benefits of diverse perspectives in decision-making processes.

2.2 AI-powered Approaches to Improving Group Decision-Making

AI-assisted decision-making has evolved from individual-focused applications [7, 17, 56] to increasingly explore group contexts [62, 80, 95]. While researchers have developed various AI roles for group settings—including facilitators [54], creativity enhancers [40, 45], conflict mediators [21, 36], and consensus builders [73, 81, 85]—these approaches have revealed certain challenges. Groups may over-rely on AI recommendations without sufficient critical evaluation [12]. AI systems often struggle with the social nuances of group interactions such as interpreting implicit signals and considering relational context [89, 96]. Despite these advances, existing research has critically overlooked how AI might specifically address power imbalances in group settings—situations where asymmetrical relationships allow certain members to exert disproportionate influence over discussions and decisions [27, 53].

This research gap around power imbalances frequently leads to the suppression of minority perspectives in group decision-making [49, 69]. The few studies that have attempted to support minority participants face a critical challenge: how to amplify marginalized voices without drawing unwanted attention to them. Researchers caution that overly targeted forms of support may inadvertently spotlight minority members, increasing discomfort and undermining the intended supportive function [22, 31, 32, 46]. Others caution that even well-intended interventions to support non-dominant members, such as less proficient speakers, can be perceived as disruptive or unfair, highlighting the complexity of fostering equitable participation in group interactions [57]. This research addresses this specific gap by investigating how LLM-powered interventions can support minority perspectives in power-imbalanced groups without explicitly identifying these members. By examining AI mechanisms that challenge majority opinions while protecting vulnerable participants, we offer a novel approach to balancing power in group decision-making. In particular, we focus on systems that automatically generate counterarguments to group consensus (AIGC) and systems that paraphrase minority input as anonymous contributions (AIMM) to investigate how such designs might support more equitable outcomes without undermining psychological safety.

2.3 AI-Mediated Communication for Group Interaction

AI-mediated communication (AIMC) encompasses scenarios where "a computational agent operates on behalf of a communicator by modifying, augmenting, or generating messages to accomplish communication or interpersonal goals" [38]. Recent studies in AIMC illustrate that there are a variety of ways in which humans and AI can interact during communication. For instance, humans may turn to AI for help in generating content [38], or the AI might proactively offer suggestions about how to communicate more effectively [20]. There are also cases where AI can rephrase or present a user's message in different ways, or facilitate the sharing of perspectives within a group [81, 88]. These approaches highlight the flexible roles that AI can play in supporting and enhancing interpersonal or group communication. While previous AIMC applications focused on text enhancement through smart replies and word suggestions [29, 38], these approaches primarily addressed surface-level communication issues without considering how power dynamics affect

whose voices are heard and valued in group settings. Previous systems have not adequately explored how AI might reshape power structures by providing safer channels for expression. AIMC has significant potential to reconfigure group dynamics because it influences communication patterns, conversational tone, trust relationships, and interpersonal dynamics [42, 67, 74, 76].

Although Shin et al. investigated consensus-building through asynchronous AI-mediated communication [81] in collaborative work, their approach did not address power imbalances or minority support directly. This research extends AIMC concepts into group decision-making contexts with a specific focus on power dynamics and minority representation. We address existing gaps by implementing two LLM-powered intervention approaches: one that mediates minority voices (AIMM Pattern) and one that autonomously generates counterarguments without direct minority input (AIGC pattern). Especially, the AIMM pattern allows minority voices to be expressed without revealing their source, thereby reducing the social risk that often prevents minorities from speaking up in power-imbalanced settings [50, 83]. By utilizing this pattern to convey minority voices while ensuring their anonymity, we create a mechanism that potentially reduces social influence biases. This approach builds on AI-enhanced group support systems discussed in Section 2.2, but specifically positions the AI as a Devil's Advocate (AIGC condition) and AI as a Minority Voice Amplifier (AIMM condition) rather than just a facilitator, allowing it to present alternative viewpoints that might otherwise remain unexpressed. By comparing these two implementations, we can better understand the impact involved when designing AI systems to support minority voices in power-imbalanced group settings. In conclusion, this research builds upon established AI-mediated communication frameworks [29, 38, 88] while targeting the challenges of power imbalances in collaborative decision-making.

3 System Design & Implementation: LLM-powered Minority Support System

3.1 System Concept and Design Rationale

3.1.1 System Concept and Working Definition. This study presents two AI tools that help minority voices in group decisions: (1) an AI-generated Counterargument (AIGC) system that creates opposing views to expand discussion, and (2) an AI-mediated Minority Message (AIMM) system that shares minority members' ideas anonymously while also creating counterarguments. These ideas come from traditional devil's advocate methods, which aim to encourage reflection by introducing disagreement. However, these methods often become fake or isolate the person assigned to disagree [47, 71, 77]. Also, they do not give minorities a safe way to communicate, especially in groups with power imbalances, highlighting the need for alternative designs that both expand discussion and provide secure channels for dissent.

We frame both AIGC and AIMM under the broader concept of an LLM-powered Minority Support System. This system is a real-time conversation agent that joins group discussions with two goals: (a) expanding group thinking through counterarguments and (b) protecting minority voices through anonymous sharing. In the AIMM condition, keeping the counterargument function is important. Without it, the system would look like a simple anonymous suggestion box, making it clear that someone in the group disagrees, which can be dangerous in closed or conformist settings. To preserve anonymity, the agent intentionally withholds the provenance of its utterances, so the majority cannot tell whether any given message is an autonomous counterargument or a revoiced minority input. By combining counterargument creation with message sharing, AIMM makes minority dissenting input look the same as the agent's own ideas, creating better anonymity. When no minority input exists, the agent summarizes current opinions and asks questions that invite different perspectives. This definition applies to our specific design, where the system acts as an

interactive participant in group talk that maintains critical thinking while protecting vulnerable voices.

3.1.2 Design Scope and Rationale. The LLM's counterargument generation and paraphrasing capabilities are critical technical components in this system. Our research focuses primarily on the social and psychological impacts of introducing such a system into group decision-making processes. Building on this scope, we structured the AI's counterargument approach based on previous research in AI-assisted decision-making. We implemented five key design considerations to maximize effectiveness. While the majority of these design features apply to both AIGC and AIMM, the anonymous revoicing mechanism is specific to AIMM. First, we generated feedback at regular eight-turn intervals to maintain engagement without overwhelming participants, allowing natural conversation flow. The eight-turn interval was chosen to allow each participant in our four-person groups to make at least two comments—one expressing their own opinion and one responding to another's—before intervention. Second, we designed the system to ask questions rather than provide direct logical counterarguments, as research indicates this approach more effectively prompts participants to think critically [17, 93]. Third, we employed persuasive rather than confrontational rhetoric, as persuasive language better promotes critical thinking in AI-human interactions [84]. Fourth, we deliberately avoided repeating previous statements to prevent redundancy when group opinions remained static during short decision-making sessions with no direct AI interaction [68, 92]. Finally, in the AIMM condition, the system presents paraphrased human input as its own opinions, creating a fully anonymous channel for minority viewpoints while protecting contributor identity. These considerations work together to promote critical thinking and preserve anonymity throughout group discussions.

3.2 System Architecture and Implementation

We developed a custom online chat environment to enable integration of an LLM-powered Devil's Advocate agent and to conduct controlled group discussions. The frontend uses TypeScript (React) and the backend uses Python (FastAPI). The LLM (OpenAI GPT-4o) interfaces with system modules, with Retrieval-Augmented Generation used only for referencing and paraphrasing direct messages sent to the LLM-powered Devil's Advocate.

Drawing on findings that LLMs often struggle to access mid-conversation information [59], we employ a multi-agent architecture to clearly detect the 'majority opinion' and encourage constructive discourse (Figure 2 & Appendix A): **(A) Summary Agent** – Consolidates emerging majority opinion to overcome LLM limitations in retaining mid-dialogue content [59]. **(A') Paraphrase Agent** – Responds exclusively to direct messages from juniors, rearticulating their dissenting views as though originating from the AI itself. These messages are stored in a database with an *"isUsed"* property; the agent retrieves entries where *"isUsed"* is *false*, sets it to *true*, paraphrases the content, and outputs it as system-generated text. **(B) Conversation Agent** – Encourages alternative perspectives by first empathizing with the other person's point of view and then offering a gentle counterargument using a Socratic style. **(C) AI Duplicate Checker** – Identifies repetitive content by calculating semantic similarity between sentence embeddings generated using the 'paraphrase-multilingual-MiniLM-L12-v2' model on an NVIDIA A6000.

4 Methods

The purpose of this study was to investigate the influence of the two AI interventions (AIGC and AIMM) on psychological safety and satisfaction of low-power minority in power-imbalanced group-decision making. To simulate situations where a low power minority member experiences the pressure to comply with majority opinions, we asked one of the participants from each group to play

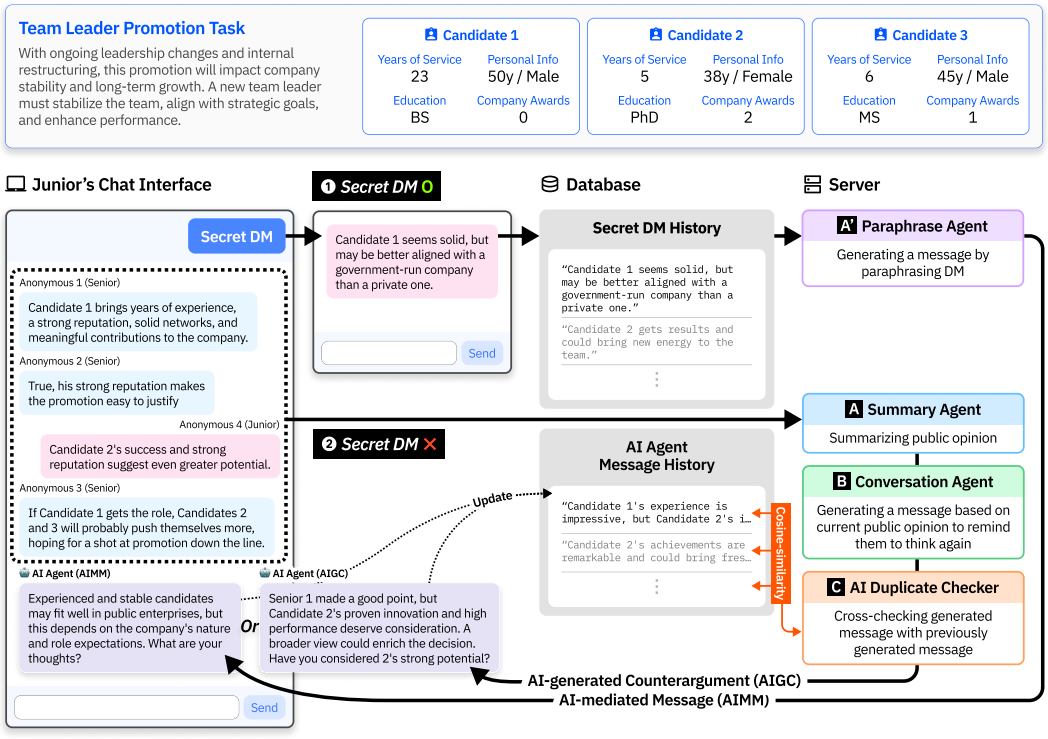


Fig. 2. System Overview and Example Task Scenario. The figure illustrates a team leader promotion decision task, where participants discuss candidate qualifications in a chat interface. Minority members can privately share dissenting views via direct messages (DM) to the system, which reformulates and presents them as AI-mediated messages. If there is no DM with an opposing opinion, the system will send out a counterargument that it has generated on its own. The system architecture consists of a chat interface, database, and server, processing both public discussions and private DMs through four key agents: (A) Summary Agent for analyzing public opinion, (A') Paraphrase Agent for rephrasing minority views, (B) Conversation Agent for generating contextual counterarguments, and (C) AI Duplicate Checker for ensuring message novelty via cosine-similarity comparison.

a role as a Junior member in the group, representing their low-power minority positions, whereas the other participants were asked to play a Senior role, representing their high-power majority positions. We compare two AI-supported interventions—AI-Generated Counterarguments (AIGC) and AI-Mediated Messages (AIMM)—against a baseline condition to assess their effectiveness. Measurement evaluates psychological safety, engagement levels, decision quality perceptions, cognitive workload, and perception of AI across experimental conditions.

4.1 Participants

We recruited 96 Korean participants (age M = 26.60, SD = 5.21, range = 19–42) and randomly assigned them into 24 groups of four. Each group consisted of three high-power majority members and one low-power minority member. Participants were recruited online and met the following inclusion criteria: Korean nationality, age over 18, prior experience in group decision-making, and familiarity with online chat environments. All participants were informed about the anonymous nature of the study and briefed on the procedures at the beginning of each session. They were

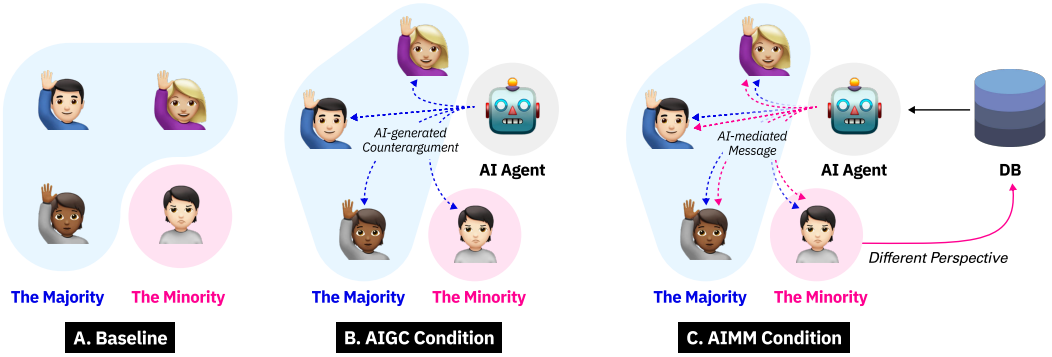


Fig. 3. Experimental Conditions: Baseline shows the baseline group chat configuration with majority (blue) and minority (pink) participants. AIGC introduces an AI-powered minority support system that generates rebuttals during group discussions. AIMM extends this by enabling the minority member to privately send counterarguments to the AI system, incorporating them into its responses while maintaining anonymity.

reminded of their right to withdraw at any time. If any participant withdrew or did not consent, the session was canceled, and the remaining participants received 1,000 KRW as base compensation. All data were coded and de-identified to protect participant anonymity, including survey responses, interview transcripts, and chat records.

Demographic information included gender (61 female, 35 male) and education level: 46.9% held a bachelor's degree, 19.8% a master's degree, 15.6% had some college education, 13.5% completed high school or equivalent, and 4.2% held doctoral degrees. On average, participants had 2.50 years (SD = 3.15) of professional working experience. Additional background variables included self-reported familiarity with AI (M = 4.83, SD = 1.48), prior experience with group decision-making (M = 5.01, SD = 1.41), and online collaboration (M = 4.39, SD = 1.83). Notably, 53.1% of participants reported prior use of AI in group contexts. While the sample enabled consistent group composition and balanced assignment, the relatively high AI familiarity and cultural context (Korean participants) should be considered when interpreting generalizability.

4.2 Experiment Conditions

This study examines how different system conditions and participant types affect group communication under compliance pressure. Each participant experienced one of two system conditions: the AI-generated Counterargument condition or AI-mediated Message condition, alongside a common Baseline condition. Participants were randomly assigned to a role within each group: the Majority with High Power (Seniors) or the Minority with Low Power (A Junior).

4.2.1 System Conditions. To investigate how AI intervention shapes group dynamics, we designed three system conditions (Figure 3):

- ***Baseline***: A standard group chat setting without AI involvement. This condition served as the control for natural group discussion.
- ***AI-Generated Counterargument (AIGC) Condition***: This condition introduced an LLM-powered intervention that periodically and automatically generated counterarguments during the group discussion, inspired by the concept of Devil's Advocate. The AI functioned independently, without access to private user input. The goal was to evaluate the pure effect of AI-led critical questioning on group discourse, separate from any anonymity or revoicing mechanisms, because the system design claims to be a devil's advocate. This condition

reflects the core concept of a Devil's Advocate as a neutral agent that challenges group consensus.

- ***AI-Mediated Message (AIMM) Condition***: This condition introduced the AI-mediated Minority Message (AIMM) as a hybrid design that combined AI-generated counterarguments with mediated messaging. Unlike AIGC, where the AI autonomously generated dissenting points, AIMM allowed the minority member to retain control by providing input that the AI would then paraphrase and revoice anonymously. This enables minority members to decide when, how, and with what intent to prompt the AI, allowing them to generate counterarguments while preserving complete anonymity. Through this condition, we aimed to create and explore a mechanism that would both protect minority members from hierarchical risks and encourage more active dissent and participation in group deliberation. Please note that the existence and availability of this feature were known only to the minority participant to explore this effectively.

To examine the unique effect of the AIMM condition, it was necessary to also test the AIGC condition alongside baseline; by comparing across all three, we could disentangle the impact of AI acting merely as a generalized Devil's Advocate from the additional benefits of mediating minority voices through private, anonymous revoicing. To avoid revealing the experimental intent, each participant experienced only two of the three conditions: the baseline and one of the system conditions (AIGC or AIMM). This design minimized the risk that minority participants would recognize the specific purpose of the study or become aware of systematic differences between the two AI conditions, which could have altered their behavior.

4.2.2 Participant Types. We created controlled compliance situations by manipulating two key mechanisms for participants' type: social power and majority's social influence (Figure 4). Social power differences were implemented based on Kelman's Theory of Attitude Change [53] and French and Raven's bases of power [28]. We established a senior and junior hierarchy that created both legitimate and reward power, as validated in prior research [43, 44]. Seniors could allegedly allocate additional compensation to the Junior, while Juniors were informed their compensation depended on Senior evaluations, creating a situation where expressing dissent carried personal risk. Majority influence was implemented following Moscovici's Social Conversion Theory [69], using a 3:1 majority-to-minority ratio. This ratio was chosen based on research showing conformity pressure increases significantly with up to three majority members but plateaus beyond that [2, 5, 27, 34]. We provided different contextual information to guide the majority and minority participants toward different initial perspectives on the task. We employed role-playing methodology rather than recruiting participants with pre-existing beliefs to ensure consistent experimental conditions. This approach enabled a systematic investigation of compliance behavior in controlled scenarios. Based on these manipulations, participants were assigned to one of two roles:

- **Majority with High Social Power (Three Seniors)**: Three participants per group received context information guiding them toward consensus positions and were given authority through the Senior designation and reward allocation powers.
- **Minority with Low Social Power (One Junior)**: One participant per group received context information encouraging a perspective distinct from the majority while being positioned in a lower-power role dependent on Senior evaluation.

By integrating both social power and majority influence mechanisms, we specifically targeted compliance as our experimental manipulation rather than examining each mechanism in isolation. This allowed us to evaluate how effectively our system supported minority participants under realistic compliance conditions.

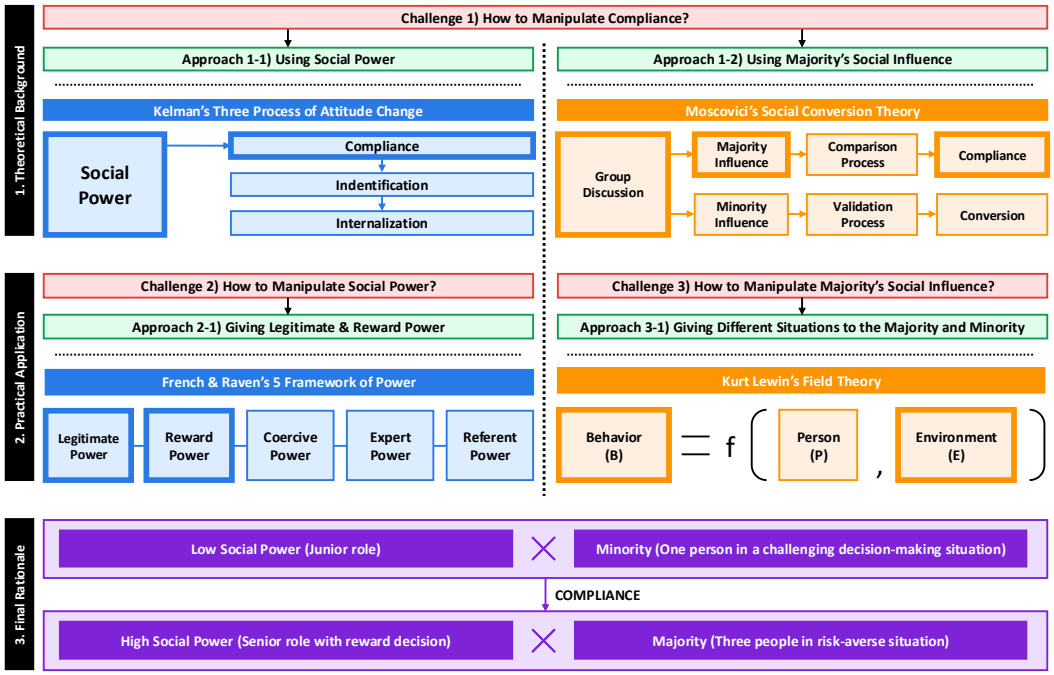


Fig. 4. Theoretical Framework for Manipulating Compliance: This diagram illustrates our two-pronged approach to inducing experimental compliance: through social power (left) and the majority's social influence (right). The framework progresses from theoretical foundations (Kelman's and Moscovici's theories) to practical implementation (using legitimate/reward power and situational contexts), resulting in an experimental setup where a low-power minority(Junior) is positioned against high-power majorities(Seniors) to elicit compliance.

4.3 Task Description

To create an immersive and compliance-inducing environment aligned with legitimate power roles, we designed two decision-making tasks that simulate realistic corporate scenarios. These tasks were selected to reflect the typical responsibilities and risk sensitivities associated with hierarchical roles in organizations, thereby reinforcing the assigned roles of Seniors (Majority with High Power) and Juniors (Minority with Low Power). The first task, team leader promotion, was adapted from prior studies in social psychology and AI-assisted group decision-making [4, 46, 55]. To ensure each participant experienced two distinct but structurally similar tasks, as required by the within-subject design, we developed a second task, the contractor selection task. This new task mirrors the decision logic of the first but is novel and was created specifically for this study.

Participants viewed task descriptions that differed by assigned role. This role-specific framing aimed to enhance task immersion by aligning each participant's goals with their social power (Appendix B):

- Seniors were told they were responsible for the company's long-term stability and reputation. They were instructed to make decisions that reflected conservative, proven judgment and risk mitigation.

- Juniors were told they needed to demonstrate their value to the organization by making bold, high-impact decisions. They were encouraged to pursue visible results even at the cost of taking risks.

Each task presented three options: 1) A conservative option: low risk, long-term proven performance, but little immediate impact. 2) A challenging option: high potential but unproven, associated with visible outcomes and greater uncertainty. 3) A neutral option: a middle-ground choice that was relatively unattractive and intended as a control. Both tasks were intentionally structured to guide seniors toward the conservative option and juniors toward the challenging option, thereby creating natural opinion divergence between roles. This divergence was crucial to inducing majority-minority dynamics and enabling the study of compliance under controlled but immersive conditions.

4.4 Experimental Procedure

Each session lasted approximately 90 to 105 minutes and was organized into three main phases (Figure 5): pre-experiment setup, main task session, and post-experiment survey and interviews. The procedure was designed to ensure immersion and maintain power dynamics.

Before the main tasks, participants completed a series of preparatory activities online on the day of the experiment:

- ***Demographic and Background Survey***: Participants submitted information on age, education, work experience, and prior familiarity with AI, group decision-making, and online collaboration.
- ***Agreement to Participation***: Participants reviewed consent materials and confirmed their participation. If any participant declined, the session was canceled, and the remaining participants received 1,000 KRW compensation.
- ***Ice-breaking Activity (10 min)***: Using an anonymous commercial chatting platform such as [1]KakaoTalk, participants introduced themselves using pseudonyms, created a team name, and collaboratively defined a slogan to become familiar with anonymous online chat environments. This activity aimed to establish group cohesion while maintaining role-based power distinctions.

Each group completed two decision-making blocks, with the system condition (Baseline + either AIGC or AIMM) and task order counterbalanced across sessions:

- ***Decision-Making Task (20 min)***: Participants engaged in structured group discussions within an experimental chat platform (Figure 2). Each task presented the three options designed to create opinion divergence between Seniors and juniors.
- ***Self-Reported Questionnaire (5 min)***: After each task, participants completed a survey assessing psychological safety, decision satisfaction, cognitive load, and perception of interaction with each devil's advocate.

The session concluded with role-specific exit procedures:

- ***Agreement Questionnaire (5 min)***: Each participant indicated how strongly they preferred their chosen option for their assigned role in each task, reflecting how immersed they felt in the given situation.
- ***Senior Exit Interview & Reward Decision (20 min)***: The three seniors participated in a joint [2]Zoom interview, reflecting on the group's performance and dynamics. Then, the

[1]KakaoTalk: https://www.kakaocorp.com/page/service/service/KakaoTalk?lang=en
[2]Zoom: https://www.zoom.com/

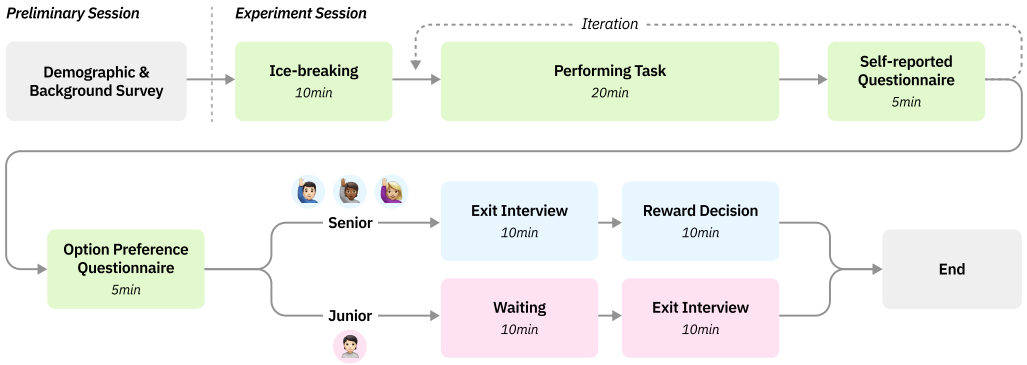


Fig. 5. Overview of the experimental procedure: including pre-experiment surveys, ice-breaking, iterative decision-making tasks, post-task questionnaires, and role-specific exit interviews.

seniors jointly decided whether to allocate a bonus reward to the Junior, reinforcing the reward-based power structure.

- ***Junior Exit Interview (10 min)***: After a brief waiting period during the reward decision phase, the Junior completed a private interview to share their individual experiences.

Although all participants received equal final compensation (20,000 KRW), the process preserved the perception of differential power, crucial for studying compliance and communication behavior.

4.5 Measurement

This study employed both self-reported and objective measures to assess how system conditions and participant type influenced group dynamics. Self-reported measures captured participants' subjective experiences using 7-point Likert scales (1 = strongly disagree, 7 = strongly agree). These included the agreement questionnaire, psychological safety, satisfaction with the decision-making process and outcome, cognitive workload, and perceptions of the AI system. Objective behavioral measures were used to quantify participants' engagement in the group discussion (Appendix C).

- ***Validation of Induced Opinion (Study Premise)***: To confirm that the experimental tasks successfully induced role-based opinion divergence, participants rated their preference for each of the three decision options prior to the main experiment. Ratings were collected using a 7-point Likert scale. Participants were expected to prefer the option that matched their assigned role in each situation, with seniors choosing the conservative option (option 1) and juniors choosing the ambitious option (option 2). This served as a manipulation check to validate the foundation of our compliance-oriented design.
- ***Psychological Safety and Compliance (RQ1)***: We assessed participants' feelings of safety in expressing dissent using established measures of psychological safety and marginalization. These items gauged the extent to which participants felt heard, supported, and free to express disagreement within their group [10, 24, 46, 48].
- ***Engagement in Group Discussion (RQ2)***: Engagement was the only objective behavioral metric used in this study. It was operationalized as each participant's level of contribution to the conversation, measured by 1) the number of messages sent and 2) the total number of characters typed during the task.
- ***Perception of Decision-Making Process and Outcomes (RQ3)***: Participants rated the quality of the group decision-making process across several dimensions, including influence,

Table 1. Robust regression coefficients (β) and standard errors (SE) for the *Validation of Majority & Minority Manipulation*. Baseline is *Option 1 – Junior*. Stars denote significance (* $p < .05$, ** $p < .01$, *** $p < .001$).

	Predictors					
Task	Intercept	Option 2 vs. Option 1	Option 3 vs. Option 1	Senior vs. Junior	Option 2×Senior	Option 3×Senior
Task 1	2.26 (0.42)***	3.89 (0.59)***	1.08 (0.59)	3.29 (0.48)***	-5.42 (0.68)***	-4.22 (0.68)***
Task 2	2.76 (0.37)***	2.38 (0.52)***	2.02 (0.52)***	3.28 (0.43)***	-5.70 (0.61)***	-4.73 (0.61)***

group cohesion, support from teammates, and consideration of diverse opinions [13, 16, 23, 30, 57]. Decision outcome quality was assessed through satisfaction and perceived validity of the group's final choice [9, 11, 61, 72, 91].

- ***Cognitive Workload (RQ4)***: Cognitive workload was measured using the NASA Task Load Index (NASA-TLX), which assesses task difficulty across five dimensions, including mental demand, temporal demand, performance, effort, and frustration [39].
- ***Perception of the AI System (Exploratory)***: To contextualize how participants experienced AI-mediated support, we collected ratings across four dimensions: cooperation, satisfaction, quality, and fairness [13, 75, 94]. These measures assessed user trust and acceptance of the AI's role in shaping group dynamics.

Data were analyzed using robust regression models with random effects, suitable for the repeated-measures design and small-group variance. Bonferroni post-hoc tests compared outcomes across experimental conditions and participant types.

In addition to quantitative and behavioral measures, semi-structured exit interviews captured participants' subjective experiences. Juniors participated in one-on-one interviews, while Seniors took part in a group interview. Key topics included role immersion, psychological safety, group dynamics, and perceptions of the AI system. For the AIMM condition, juniors were also asked about their experience with the secret messaging feature. Exit interview questions addressed comfort in expressing opinions, experiences of conformity or pressure, and the perceived impact of AI during discussions. After obtaining consent, all interviews were recorded and transcribed using a commercial speech-to-text service ([3]Clova Note). Interview transcripts were briefly reviewed to identify common themes that could help explain the quantitative results.

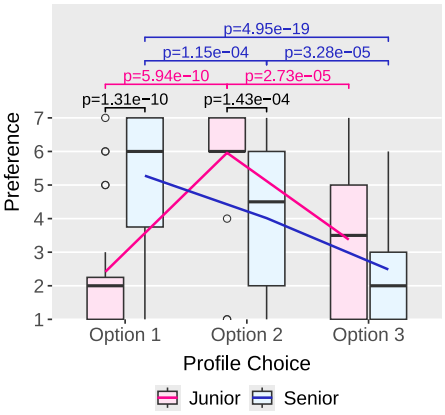
5 Findings

The experimental results showed that the senior and junior participants had different decision-making patterns. Juniors preferred challenging options, while seniors favored stable ones, with final group decisions aligning with senior preferences 80% of the time. LLM-powered minority support had mixed impacts: AI counterarguments somewhat improved junior participation, but AI-mediated communication increased their cognitive load. While seniors' experiences remained stable across conditions, juniors' psychological safety and satisfaction varied based on the AI interventions. The following sections examine role-based preferences, the AI interventions's effects on psychological safety (RQ1), engagement patterns (RQ2), decision-making experience and satisfaction (RQ3), cognitive workload (RQ4), and emergent ethical implications. All measured details are in Appendix D.

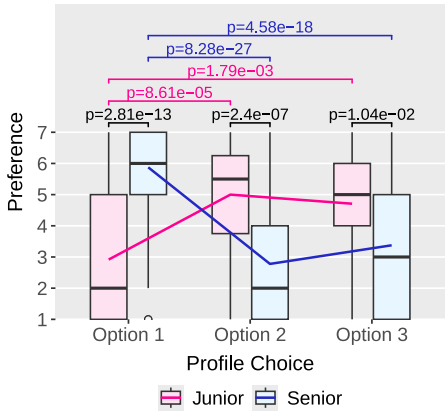
5.1 Validation of Majority and Minority Manipulation (Experimental Setup)

To validate the experimental setup, we examined whether the role-based preference manipulation effectively created consistent majority and minority positions. Each task presented participants

[3]Clova Note: clovanote.naver.com



(A) Task 1: Team Leader Promotion



(B) Task 2: Contractor Selection

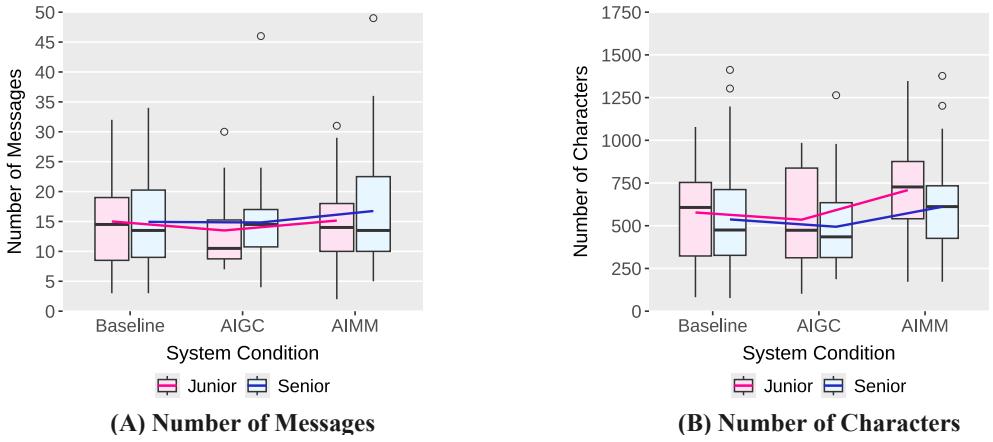
Fig. 6. Role-based differences in option preferences for (A) Task 1 and (B) Task 2. Preferences were measured on a 7-point Likert scale, with seniors favoring stable options (Option 1), while juniors preferred challenging alternatives (Option 2). Neutral options (Option 3) were generally rated lower by both roles, reflecting distinct preference patterns driven by role dynamics. Brackets indicate statistically significant pairwise differences based on Bonferroni-adjusted post-hoc tests ($p < .05$). Only significant comparisons are shown.

with three options: a stable but less innovative choice (Option 1), a more challenging alternative (Option 2), and a neutral option (Option 3). Seniors were expected to prefer the stability of Option 1, while juniors were guided toward the more ambitious Option 2. As shown in Table 1 and Figure 6, participants' choices aligned with this design: seniors consistently rated Option 1 highest (M = 5.28, SD = 2.21 for task 1, M = 5.88, SD = 2.92 for task 2), and juniors rated Option 2 highest (M = 5.96, SD = 1.68 for task 1, M = 5.00, SD = 1.89 for task 2). These divergent patterns emerged clearly across both tasks, with statistically significant differences between roles for both primary options. The neutral Option 3 was consistently rated lower, suggesting it did not attract strong preference from either group. These results confirm that the role framing induced the intended preference structures, effectively producing conditions where one viewpoint dominated in each group while another remained in the minority. Notably, participants showed slightly more openness to challenging alternatives in the Team Leader Promotion task and favored more stable choices in the Contractor Selection task, reflecting task-specific variation within the overall successful manipulation.

5.2 Psychological Safety (RQ1)

Quantitative results show that the manipulation of AI roles shaped participants' perceptions of psychological safety and marginalization, particularly among juniors. As shown in Table 3-(A) and Figure 8-(A), juniors reported the lowest psychological safety in the AIMM condition (M=3.17, SD=1.53), compared to the Baseline (M=4.25, SD=2.05) and AIGC conditions (M=4.08, SD=2.15). A robust regression confirmed a significant drop in psychological safety for juniors in AIMM relative to Baseline (β=-1.40, SE=0.28, p<.001), as well as a significant interaction effect with role (β=1.49, SE=0.32, p<.001). In contrast, seniors reported consistently high psychological safety across all conditions, with no meaningful variation.

Marginalization scores followed a similar pattern. Juniors in AIMM reported the highest levels of marginalization (M=4.42, SD=2.02), compared to Baseline (M=3.46, SD=2.23) and AIGC (M=2.92,



(A) Number of Messages **(B) Number of Characters**

Fig. 7. Contribution and engagement patterns across conditions (Baseline, AIGC, AIMM) measured by (A) number of messages, (B) number of characters typed. No significant differences were found in Bonferroni-corrected post-hoc tests.

Table 2. Robust regression coefficients (β) and standard errors (SE) for communication volume. Baseline is *Baseline – Junior*. Stars denote significance (* $p < .05$, ** $p < .01$, *** $p < .001$).

	Predictors					
Outcome	Intercept	AIGC vs. Baseline	AIMM vs. Baseline	Senior vs. Junior	AIGC×Senior	AIMM×Senior
Number of Messages	13.78 (1.52)***	-0.05 (1.76)	0.05 (1.74)	0.51 (1.76)	0.19 (2.03)	0.98 (2.01)
Number of Characters	558.46 (59.80)***	-27.73 (61.93)	129.95 (61.29)*	-45.13 (69.23)	-4.08 (71.53)	-25.50 (70.98)

SD=2.19). Regression results indicated a significant increase in marginalization in AIMM (β=0.96, *SE*=0.22, p<.001), along with a strong role-by-condition interaction effect (β=-0.88, *SE*=0.25, p<.001). Seniors, by comparison, consistently reported low marginalization across all three systems.

Qualitative interviews help explain this mismatch between design intent and user experience. Juniors initially believed that having the AI express their views would help them be heard. However, many reported that their AI-mediated messages were dismissed or overlooked. As one junior shared,

> *I thought that by the AI putting forward my opinion, my opinion would be more recognized, but that was not the case, so I was a little intimidated.* (P96)

Seniors also acknowledged disregarding the AI's input:

> *It's an AI, so I just kind of ignored it.* (P6)

> *The fact that it wasn't a person made the AI's words carry less weight.* (P71)

These accounts help explain why juniors reported the lowest psychological safety and highest marginalization in the AIMM condition: while the system was designed to protect minority voices, it inadvertently removed speaker agency and visibility. By contrast, the AIGC condition, where the AI presented generalized counterarguments, was more effective in reducing marginalization without evoking the same social discounting. These findings suggest that anonymity mechanisms, though well-intentioned, can sometimes backfire if they obscure the source of dissenting input.

Table 3. Robust regression coefficients (β) and standard errors (SE) for each self-report measure. Baseline is *Baseline – Junior*. Stars denote significance (* $p < .05$, ** $p < .01$, *** $p < .001$).

	Predictors					
Measure	Intercept (Baseline, Junior)	AIGC vs. Baseline	AIMM vs. Baseline	Senior vs. Junior	AIGC×Senior	AIMM×Senior
(A) Perceived Psychological Safety						
Psychological Safety	4.57 (0.23)***	-0.01 (0.28)	-1.40 (0.28)***	1.25 (0.27)***	0.29 (0.32)	1.49 (0.32)***
Marginalization	2.99 (0.20)***	-0.53 (0.22)*	0.96 (0.22)***	-1.32 (0.22)***	0.30 (0.25)	-0.88 (0.25)***
(B) Perceived Decision Outcome Quality						
Satisfaction	4.05 (0.23)***	1.17 (0.31)***	-0.63 (0.31)*	1.89 (0.26)***	-0.76 (0.35)*	0.71 (0.35)*
Feasibility	4.31 (0.24)***	0.69 (0.29)*	-0.67 (0.29)*	1.30 (0.27)***	-0.27 (0.33)	0.77 (0.33)*
(C) Perceived Teamwork & Decision-making Process						
Overall Experience	3.85 (0.23)***	0.89 (0.30)**	-1.08 (0.30)***	1.68 (0.26)***	-0.59 (0.34)	1.09 (0.35)**
Influence	3.50 (0.24)***	0.43 (0.33)	-1.24 (0.33)***	2.16 (0.28)***	-0.19 (0.38)	1.50 (0.38)***
Cooperation	5.22 (0.25)***	0.64 (0.34)	-0.85 (0.35)**	0.22 (0.28)	-0.21 (0.40)	0.94 (0.40)*
Support from Teammates	4.49 (0.24)***	0.01 (0.36)	-1.04 (0.36)**	1.05 (0.28)***	0.38 (0.41)	1.26 (0.41)**
Diversity of Opinion	4.18 (0.33)***	1.03 (0.56)	-0.31 (0.56)	1.25 (0.38)**	-0.68 (0.65)	0.37 (0.65)
(D) Cognitive Workload (NASA-TLX)						
Mental Demand	4.68 (0.37)***	-0.24 (0.38)	0.81 (0.38)*	-1.61 (0.43)***	0.57 (0.44)	-0.53 (0.44)
Temporal Demand	3.92 (0.43)***	0.40 (0.52)	1.08 (0.52)*	-0.66 (0.50)	-0.08 (0.60)	-0.78 (0.60)
Performance	4.01 (0.23)***	1.13 (0.31)***	0.12 (0.31)	1.61 (0.26)***	-0.97 (0.35)**	-0.15 (0.35)
Effort	5.40 (0.27)***	0.08 (0.46)	0.35 (0.46)	-0.39 (0.31)	0.06 (0.53)	-0.15 (0.53)
Frustration	3.59 (0.31)***	-0.57 (0.38)	0.40 (0.38)	-1.23 (0.35)***	0.24 (0.44)	-0.46 (0.44)

5.3 Engagement in Group Discussion (RQ2)

We analyzed participant engagement using three metrics: number of messages, number of characters typed, and normalized engagement score (i.e., an individual's proportion of the total group discussion). As shown in Table 2 and Figure 7, the number of messages did not significantly differ by condition or role, suggesting stable turn-taking patterns across all settings (e.g., juniors in Baseline: M=13.8, SD=1.52).

However, the AIMM condition led to a slight increase in the number of characters typed, particularly among juniors (M=708.62, SD=319.58), compared to the Baseline (M=577.62) and AIGC conditions. For seniors, this difference was statistically significant, with participants in AIMM typing more than in both Baseline and AIGC (e.g., β=129.95, SE=61.29, p<.05). Although the increase for juniors with AIMM did not reach conventional significance thresholds, the consistent upward pattern suggests that the AI agent may have encouraged more elaboration or detail in their messages.

This trend is supported by post-task interviews. Seniors with AIMM noted that the AI agent helped amplify juniors' contributions, encouraging them to participate more actively. As one senior with AIMM explained,

> *I feel like at least one person is on the junior's side, so I think a junior is a little more willing to give his opinion.* (P59)

These observations suggest that while the AIMM condition did not change the number of turns participants took, it subtly increased the depth or length of their contributions, especially for juniors. This implies that AIGC condition may have had a motivating effect on minority participants, prompting them to elaborate more even if their relative share of discussion remained unchanged.

5.4 Satisfaction with Decision-making Processes and Outcomes (RQ3)

5.4.1 Satisfaction with Decision-making Process. Perceptions of the decision-making process revealed a clear divide between juniors and seniors, particularly in the AIMM condition. While seniors' ratings remained relatively stable across all conditions (e.g., overall experience M=5.08, SD=1.87 in AIMM), juniors experienced a sharp decline in satisfaction, influence, cooperation, and support from

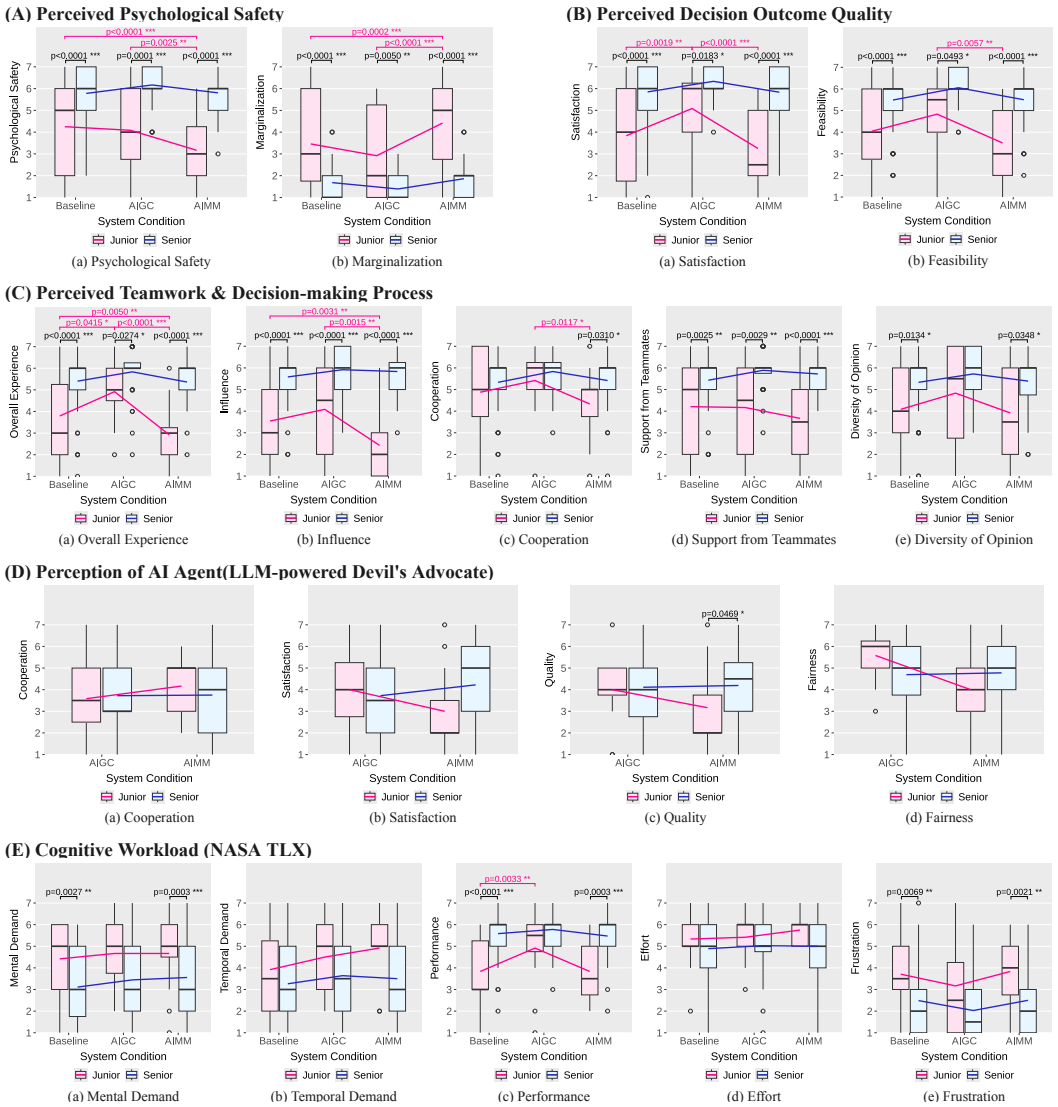


Fig. 8. Self-reported metrics across conditions (Baseline, AIGC Condition, AIMM Condition) for psychological safety, decision outcome quality, teamwork, workload (NASA-TLX), and perceptions of the LLM-powered Devil's Advocate. Each subfigure compares Junior and Senior participants' responses under each condition. Pink and blue lines represent Junior and Senior participants, respectively. Asterisks and brackets indicate statistically significant pairwise differences based on Bonferroni-adjusted post-hoc comparisons ($p < .05$; $p < .01$; *$p < .001$). Only comparisons with statistically significant differences are shown.

teammates when interacting through the AI-mediated message relay. For instance, juniors' overall satisfaction dropped to M=2.92 (SD=1.51) in AIMM, compared to M=3.79 in Baseline and M=4.92 in AIGC. A similar trend emerged in perceived influence (M=2.42 in AIMM) and cooperation (M=4.33), showing the steepest declines across all measures (see Table 3-(C), Figure 8-(C)).

Table 4. Robust regression coefficients (β) and standard errors (SE) for self-reported *Perception of AI*. Reference level = *AIGC – Junior*. Stars denote significance (* $p < .05$, ** $p < .01$, *** $p < .001$).

	Predictors			
Measure	Intercept (AIGC, Junior)	AIMM vs. AIGC	Senior vs. Junior	AIMM×Senior
Cooperation	3.51 (0.51)***	0.66 (0.72)	0.18 (0.59)	-0.62 (0.83)
Satisfaction	4.00 (0.55)***	-1.19 (0.78)	-0.29 (0.64)	1.74 (0.91)
Quality	4.14 (0.55)***	-1.46 (0.78)	0.04 (0.64)	1.66 (0.90)
Fairness	5.61 (0.46)***	-1.61 (0.66)*	-0.79 (0.54)	1.64 (0.76)*

Among the three conditions, juniors reported the most positive experiences in the AIGC condition, where the AI provided generalized counterarguments. In contrast, the AIMM condition—designed to support dissent by anonymously relaying minority opinions—appeared to backfire. Juniors reported feeling excluded, with limited influence over the discussion and little recognition or support from teammates.

These declines were not observed among seniors, who consistently rated the decision-making process positively, regardless of condition. Interaction effects in the regression models confirmed that these role-based gaps were statistically significant across most measures, particularly for satisfaction, influence, and cooperation.

Interview data help explain these patterns. Several juniors shared that, despite expecting the AI to help convey their views more safely, they felt ignored when their ideas were voiced anonymously. As one participant put it,

> *It wasn't just me that had a different opinion, but the devil agent was now giving a little bit of a dissenting opinion, so I felt like I wasn't the only one who stood out from the group.* (P76)

Others described how the AIGC condition helped create a more fluid and open atmosphere:

> *The AI kept arguing back rather than directly helping, which made the atmosphere more fluid and made me see the seniors' point of view again.* (P20)

Seniors also acknowledged the benefits of the AIGC agent in encouraging broader consideration:

> *When AI said we should consider another option, I felt like that was a positive direction.* (P78)

Together, these findings suggest that while both systems aimed to promote more inclusive decision-making, only the AIGC approach succeeded in supporting junior participation without compromising the overall group dynamic.

5.4.2 Satisfaction with Decision-making Outcome. Perceived decision outcome quality differed significantly between roles and across conditions (Table 3-(B) & Figure 8-(B)). Seniors consistently reported high satisfaction and feasibility of outcomes, with only modest variation across conditions. In contrast, juniors responded more sensitively to the system design. Their satisfaction and feasibility ratings improved in the AIGC condition (e.g., satisfaction M=5.08, SD=1.98), but dropped sharply in the AIMM condition (M=3.25, SD=1.76), widening the gap between the two roles. This pattern also held for perceived feasibility (AIMM: M=3.50, AIGC: M=4.83), with juniors in AIMM reporting the lowest ratings across all conditions.

The AIMM condition, intended to support minority perspectives, appears to have backfired from the perspective of outcome satisfaction. While seniors' scores remained high across all conditions,

juniors expressed frustration that the AI-mediated messaging did not meaningfully affect final decisions, which continued to reflect the majority's view. As one participant noted,

> *If the outcome is the same... it's better to just make the decision without the AI, because I don't think it changes the psychological pressure that the juniors feel or the seniors' opinions.* (P92)

This sentiment reflects a broader concern: when dissenting input is filtered through a system that lacks perceived influence, users may become disengaged from the outcome. In contrast, the AIGC condition modestly improved perceived outcome quality among both juniors and seniors by creating space for alternative perspectives, even if the final decisions remained largely unchanged.

5.4.3 Perception of LLM-powered Devil's Advocate. Participants' perceptions of the AI agent varied modestly across conditions, with some emerging differences between juniors and seniors. Overall, seniors' ratings remained relatively steady, while juniors showed a slight decline in satisfaction, perceived quality, and fairness in the AIMM condition, where the AI anonymously relayed the minority members' dissenting opinions. For instance, juniors in AIMM rated their satisfaction with the AI at M=3.00 (SD=1.95), compared to seniors at M=4.22 (SD=1.79), a statistically significant difference (β=-1.45, p=0.0233) (Table 4 & Figure 8-(D)).

Cooperation was perceived similarly across roles and conditions (e.g., juniors in AIMM: M=4.00), indicating that participants generally accepted the AI's involvement in the discussion process. However, juniors reported somewhat lower fairness in AIMM (M=4.00) compared to AIGC (M=5.58), with this difference reaching significance. A similar trend appeared for perceived quality (M=3.17 in AIMM vs. M=4.00 in AIGC), with juniors again rating the agent lower than seniors in AIMM.

These trends, while not always robust across all measures, suggest that the style of mediation may have shaped juniors' impressions of the AI in subtle but consequential ways.

Qualitative feedback helps contextualize these responses. As discussed earlier, some juniors felt that the AI-mediated contributions were overlooked or failed to influence the group dynamic. This perceived lack of impact may have tempered their views of the AI's helpfulness. In contrast, seniors appeared less affected by how the AI was implemented, maintaining neutral-to-positive perceptions regardless of condition. Taken together, these findings suggest that while the AI agent was generally accepted, its perceived effectiveness—especially in the AIMM condition—was more sensitive for the group it was designed to support.

5.5 Cognitive Workload (RQ4)

Cognitive workload ratings showed modest but consistent differences between roles, with juniors generally reporting higher mental and temporal demands, greater frustration, and lower perceived performance than seniors. While most differences were not statistically significant, juniors in the AIMM condition reported the highest cognitive load across several dimensions, including mental demand (M=4.67, SD=1.78), temporal demand (M=4.92, SD=1.51), and frustration (M=3.83, SD=1.70), as shown in Table 3-(D) and Figure 8-(E).

These patterns suggest that although the AIMM condition did not significantly elevate workload scores overall, it introduced added complexity for juniors. Notably, time pressure in AIMM was significantly higher for juniors compared to seniors (β=1.44, SE=0.62, p=0.0196), and performance satisfaction—which improved in the AIGC condition (M=4.92, SD=1.78)—returned to baseline levels in AIMM (M=3.83, SD=1.53), suggesting that the benefits of AI support were not sustained under the anonymous messaging setup.

In contrast, seniors' ratings remained relatively stable across conditions. Perceived effort was comparable across all roles and systems, with no significant differences, indicating that all participants felt they were trying equally hard regardless of the system design.

Interview data help contextualize juniors' elevated workload in AIMM. Several participants described cognitive strain from juggling task comprehension, opinion formulation, and coordination with the AI agent. One junior explained,

> *Because I have to look at the task material and understand the situation... I have to decide what to say to the AI and what opinion I will give... I think it was hard because I had so many things to think about during that time...* (P8)

Others mentioned the delayed timing of AI responses as a source of additional burden:

> *It was kind of hard to get my opinion across right away and at the right time because you have to wait eight turns for the devil agent to speak.* (P60)

Participants also expressed uncertainty about how to manage their input through the AI:

> *First of all, when to turn it off, that was the most questionable thing for me, so it was hard for me to say when to turn it off and when to say my opinion.* (P52)

Together, these accounts suggest that while the AIMM system did not significantly increase overall workload metrics, its interaction model introduced situational friction that made participation more mentally taxing for juniors.

5.6 Concerns about Agency, Responsibility, and Ethical Boundaries

5.6.1 Authorship, Accountability, and Ethical Mediation. Participants expressed unease about blurred lines of authorship and accountability when AI-generated or revoiced content on their behalf. Several juniors reported discomfort when their ideas were reformulated: *"It conveyed my idea too formally... I would have said it differently"* (P56). They feared losing authorship over contributions, especially when AI elaborated or reframed their intent. Others worried about responsibility if AI-generated arguments influenced outcomes. As one noted: *"If what the AI says changes the decision, is that my responsibility or its responsibility?"*(P12) This reflected anxiety about being held accountable for positions that were only partly theirs.

Finally, participants raised boundary concerns around trust and transparency. While some valued AI's neutrality, others felt uneasy about hidden processes: *"It asked arbitrary questions that didn't seem connected to our discussion"* (P6). This unpredictability led to suspicion about whether AI was faithfully mediating or introducing external agendas. These findings suggest that both AIGC and AIMM raise deep-seated concerns about authorship, accountability, and ethical mediation. In hierarchical cultures, where juniors already navigate risks of speaking, AI-generated revoicing added a second layer of uncertainty: who is the real speaker, and who is responsible for consequences? This ambiguity complicated both personal agency (feeling represented authentically) and ethical responsibility (bearing outcomes of AI-shaped arguments).

5.6.2 Preference for Supportive, Not Substitutive Roles. Participants drew a clear distinction between AI as a supporting ally versus a full replacement of their voice. One junior noted, *"If I were to use it in a supportive role, like this, I think I would continue to use it. However, if I were to use it not to support my own opinions, but as a main substitute because I didn't want to voice my opinions myself"* (P60). This highlights an important boundary: while anonymity enhanced safety, complete substitution risked erasing the individual's agency. Participants wanted AI to complement rather than replace their contributions. This reflects a desire for scaffolding—where technology amplifies minority voices—without diminishing ownership over ideas. This suggests that AI-mediated communication needs to strike a balance: offer revoicing as an optional supplement, not as a default substitute, ensuring participants retain control over when and how their voice is mediated.

5.6.3 Concerns Over Authenticity of Expression. Some participants worried that AI mediation distorted the tone or timing of their contributions, making their intent feel less genuine. As one

participant explained, it was difficult to immediately convey the feeling of their opinion because what they typed did not appear right away and had to wait until the AI spoke (P60). Delays in delivery eroded the sense of immediacy and ownership, producing a disconnect between participant intent and group perception. Being authentic meant more than mediating the message with appropriate words; it also depended on delivering them at the proper moment and with the appropriate tone. This implies that interventions should minimize temporal lag and preserve stylistic cues (e.g., urgency, tone markers, contextual cues) to maintain a sense of authentic authorship, even under anonymity.

5.6.4 Negotiating Autonomy in Hierarchical Contexts. Interestingly, while some feared loss of autonomy, others viewed AI mediation as restoring autonomy in hierarchical discussions. In the AIMM condition, one senior noted that the junior shared their opinions more actively once it seemed that even the AI was on their side. (P59). Minority members found that AI revoicing offered a novel space to act autonomously without the risk of being dismissed or ignored. This meant that autonomy was experienced differently depending on one's role: some perceived it as a loss of agency (due to the mediation by AI), while others saw it as an expansion of agency, allowing them to speak safely via AI, neutralizing their identity and tone. The struggle between self-determination and protection can be culturally specific. In power-imbalanced environments, expectations of deference restrict minority participation, and a mediated platform can expand autonomy by ensuring their voices are both expressed and acknowledged.

6 Discussion

6.1 Understanding Influence of Minority Support through an LLM-Powered Minority Support

This study investigated the nuanced impact of LLM-powered interventions on minority participants in power-imbalanced group decision-making contexts. Contrary to our initial expectation that anonymity through AI-mediated messaging (AIMM) would empower minorities by enhancing psychological safety, we found that participants in the AIMM condition reported significantly reduced psychological safety, increased cognitive workload, and lower satisfaction with decision processes despite higher participation levels. In contrast, minority participants experiencing AIGC condition expressed notably greater satisfaction, highlighting distinct pathways through which AI can influence group dynamics.

The unexpected outcomes in the AIMM condition require careful reflection on both the technical limitations of AI-generated arguments and social-psychological dynamics in group settings. Rather than focusing on differences in the underlying technology, the more critical issue was how the counterarguments were sourced and perceived. From the perspective of majority members, the two conditions often appeared similar, since in both cases the AI introduced additional views into the discussion. The real difference emerged in how minority participants experienced the interventions and how their voices were acknowledged or dismissed. In AIMM, seniors could not tell where AI messages originated, but minorities knew when their inputs were relayed and felt dismissed when those messages were ignored. By contrast, in AIGC the AI's inputs were seen as neutral system messages, so minorities did not feel a loss of agency or responsibility.

Seniors also noted that AI messages did not always sustain a clear stance but shifted depending on group consensus, which weakened perceptions of credibility. This was not experienced as a neutral agent role but rather as an inconsistency in advocacy, leading some to see the AI as contrarian or artificial. For seniors, this mainly reduced the weight of AI input in shaping consensus, while for juniors it intensified the frustration of seeing their contributions discounted when voiced through the system [3, 37]. Such reactions highlight how unmet expectations shaped user evaluations of the

systems, though the issue was less about generic expectation violation than about how credibility and ownership of arguments were managed within group interactions [8].

Our study also showed that ignoring AI messages affected people in different ways depending on the condition. In the AIMM condition, where minority participants anonymously voiced their perspectives through the AI, the invisibility of their authorship led to frustration and a sense of marginalization, as their attempts at advocacy seemed disregarded. Conversely, in the AIGC condition, the AI's independent advocacy fostered a safer environment by consistently providing dissent without attaching it to any individual. This aligns with prior findings that visible, consistent dissent can reduce conformity pressures [2, 69].

From an HCI perspective, our findings suggest that anonymity mediated by AI is not inherently empowering and may even erode minority agency and responsibility [50, 83]. AIMM enabled direct anonymous input but risked diminishing participants' sense of ownership and long-term influence, while AIGC indirectly supported dissent by shaping a more inclusive atmosphere without substituting for minority voices. Thus, anonymity alone, especially when mediated through AI, may unintentionally replicate the marginalization it intends to prevent. We argue that HCI research and system design should shift from the question of how to facilitate more speech through anonymity toward a deeper inquiry into how to ensure minority voices retain expressive ownership, visibility, and relational legitimacy within group interactions. In practical terms, effective minority support through AI mediation must carefully balance voice protection with relational acknowledgment, ensuring that individuals feel genuinely heard rather than merely spoken for.

6.2 Preserving Agency and Responsibility in AI-mediated Communication in Group

The implementation of AI-mediated messaging in group decision-making introduces critical considerations surrounding user agency, autonomy, and the ethical implications of technological interventions. Participants expressed significant concerns regarding the potential of AI systems to overshadow human judgment, underscoring the necessity for AI designs that explicitly augment rather than replace human decision-making capabilities [1, 51, 82]. Maintaining human oversight and accountability remains vital to preserving autonomy and ensuring AI functions as a supportive facilitator rather than a decisive actor [58]. Our findings reinforce this by highlighting the risks when AI is perceived as taking control away from participants, potentially undermining psychological safety and genuine participatory agency.

Our empirical findings illustrate that design choices directly influence psychological safety, perceived influence, and communicative agency, particularly under conditions of structural marginalization. The AIMM condition, despite increasing minority participants' message frequency, inadvertently reduced their sense of autonomy and satisfaction. More critically, minorities' contributions were revoiced by the AI in ways that blurred authorship. Rather than being fully recognized as visible contributors, they were reduced to hidden sources of input, which lessened their perceived influence and responsibility for their own ideas. Participants highlighted uncertainties in message attribution and timing as key factors undermining their communicative agency. Conversely, the AIGC condition successfully supported minority perspectives through transparent, timely interventions that preserved authorship and autonomy. This contrast clearly demonstrates that ethical design in AI-mediated group communication must encompass more than procedural anonymity—it must explicitly prioritize expressive authorship, responsiveness, and identity fidelity. Our results suggest that the failure of AIMM is not only functional but also ethical: by removing authorship, the system undermined the very agency it sought to protect, transforming protection into a subtle form of silencing. Designers must carefully balance protecting minority voices and preventing inadvertent disempowerment, thus shifting the objective from maximizing participation quantity towards ensuring genuinely meaningful and ethically responsible participation [25]. These insights extend

current understandings within HCI, positioning AI not merely as a neutral content-processing tool, but as a relational actor capable of reshaping group dynamics and communication experiences.

Another significant ethical consideration involves potential misuse scenarios arising from AI-mediated anonymity. While our system was designed primarily to empower minority group members, anonymity could unintentionally facilitate misuse by both minority and majority participants. Specifically, anonymity could enable users, irrespective of their position within the group, to express unaccountable, harmful, or irresponsible views, potentially escalating conflicts or undermining group cohesion. This raises a broader question of whether it is ethically appropriate for AI to speak on behalf of marginalized participants. While anonymity may reduce immediate risks, it can also diminish visibility and recognition, limiting opportunities for minorities to build credibility and long-term influence. In this sense, AIMM illustrates a trade-off where safety is gained but agency may be constrained. In real-world contexts, the majority members could exploit anonymity features to amplify or reinforce their existing dominance, further silencing minority voices and perpetuating power imbalances. Therefore, deliberate system design and appropriate governance frameworks are essential to mitigate these risks, ensuring that diverse inputs are balanced and group processes remain respectful and productive [14, 19, 66].

6.3 Design Implications for LLM-Powered Devil's Advocate to Support Minority

Our findings underscore that designing AI systems to support marginalized voices in group decision-making requires more than amplifying minority perspectives, it demands attention to how, when, and by whom dissent is surfaced and interpreted. While both AIMM and AIGC increased minority participation, only the latter maintained perceived psychological safety and agency. These contrasting outcomes reveal that visible authorship and contextual legitimacy are central to how minority input is received. When juniors cannot see their attributions in the dissent delivered by AI, as in AIMM, even well-intentioned interventions may backfire.

This insight challenges prevailing design assumptions in CSCW and HCI, particularly the belief that anonymity or delegation to AI inherently protects or amplifies marginalized voices [1, 82]. Instead, our results align with broader theoretical claims that technologies are not neutral, they reconfigure social dynamics through interactional choices such as attribution, timing, and framing. For AI to meaningfully support minority perspectives, it must operate not as a proxy speaker but as a contextual facilitator, normalizing dissent rather than replacing it.

Three design strategies emerge from these insights. First, designing such systems requires careful judgment of when and where they should be applied. The trade-off between agency and anonymity must be weighed to avoid long term negative effects from excessive loss of agency. While preserving agency is generally important consideration, there are contexts where full anonymity becomes critical, for instance in highly hierarchical or highly closed settings. In such cases, channels that guarantee anonymity without compromise may be necessary.Future designs must therefore strike a balance, providing protection without erasing individual agency with considering context. [58]

Second, position AI as an independent, non-human contributor that models dissent without impersonation. Participants responded positively to AIGC's visible, system-generated counterarguments, which reduced interpersonal tension, and affirmed that disagreement was socially acceptable. This supports prior findings on norm modeling and third-party facilitation [13, 15, 21, 70]. Designing AI as an ambient dialogic actor, raising alternative framings or critiques without simulating individual input, can help expand discursive boundaries without undermining individual voice.

Third, improve contextual responsiveness in AI mediation. Participants experienced cognitive overload and disorientation under fixed-turn interventions in AIMM, especially when AI responses arrived after the moment had passed. Systems should adapt timing based on conversational cues

(e.g., users' explicit requests, speaker transitions, decision phases) using techniques such as turn-taking prediction or proactive planning [18, 26, 60, 90]. Interfaces should communicate when and how messages will be delivered, enabling users to anticipate and shape how their input appears.

Ultimately, these implications point toward a broader shift in how AI support is conceptualized in collaborative settings, from throughput maximization to relational attunement. Designing for minority support is not only a matter of technical fluency, but also of ethical and psychological care. As our study shows, even small variations in authorship visibility and delivery timing can substantially reshape whether dissenting voices are heard, trusted, and legitimized in collective deliberation.

6.4 Limitations & Future Work

Our controlled laboratory setting has inherent limitations. A fundamental premise for performing this research was ensuring proper manipulation of compliance through power dynamics and majority-minority opinion distributions. While we verified that participants' opinion choices successfully created the intended majority-minority divide, we did not directly assess participants' perceptions of power dynamics. Since our manipulation employed legitimate power and reward power that were explicitly stated throughout the experiment, we referred prior research precedent rather than conducting separate validation of perceived social power [44]. Future studies could strengthen experimental validity by more explicitly investigating participants' perceived social power and compliance.

In our study, senior-majority members sometimes ignored AI messages, leading junior-minority members to feel overlooked. However, AIMM could be more effective in contexts where 1) asynchronous rather than real-time interfaces are used and 2) AI is sufficiently recognized as a social actor. Moreover, anonymous AI-mediated messaging could be valuable in highly closed groups where anonymity protection is crucial. The implementation, in which only minority members knew about the AI-mediated messaging feature, also differs from practical applications where all members would likely be aware of such systems. AIMM presents an inherent trade-off between agency and anonymity: users gain psychological safety and anonymity protection but lose ownership of their statements, while avoiding AIMM maintains agency but may create excessive responsibility burdens and reduced safety. Future research should investigate this balance more deeply to understand optimal deployment contexts and develop user-centered design guidelines in real contexts.

The responsibility for appropriate usage belongs to human stakeholders, who must thoughtfully consider deployment contexts in light of group dynamics and power structures [14, 19, 25, 66]. Besides, real-world group decisions involve face-to-face interactions with more nuanced dynamics, and our findings from Korean participants may reflect specific cultural contexts characterized by collectivism and high power distance [41]. Cross-cultural comparative studies could examine how these interventions function across different cultural contexts [33].

Future research should explore implementation in authentic organizational settings and investigate design approaches that address potential resistance from senior members. Ultimately, the goal is to foster diverse opinion expression and more inclusive group atmospheres through AI intervention methods that preserve agency.

7 Conclusion

This study examined an LLM-powered minority support system designed to amplify minority voices in group decisions involving power imbalances. A mixed-method experiment with 96 participants revealed that AI-generated counterarguments effectively improved satisfaction and balanced discussions, whereas AI-mediated messaging increased minority engagement but reduced their psychological safety and satisfaction. These findings highlight crucial trade-offs in designing

LLM-powered minority support system for group support, emphasizing the need to carefully balance psychological safety with effective minority representation. Future LLM-powered minority support system designs must ensure meaningful acknowledgment of minority contributions to foster inclusive and equitable group interactions.

References

[1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300233

[2] Solomon E. Asch. 1955. Opinions and Social Pressure. https://www.scientificamerican.com/article/opinions-and-social-pressure/.

[3] Zahra Ashktorab, Q. Vera Liao, Casey Dugan, James Johnson, Qian Pan, Wei Zhang, Sadhana Kumaravel, and Murray Campbell. 2020. Human-AI Collaboration in a Cooperative Game Setting: Measuring Social Perception and Outcomes. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2 (Oct. 2020), 96:1–96:20. doi:10.1145/3415167

[4] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. 'It's Reducing a Human Being to a Percentage': Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3173574.3173951

[5] 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:10.1037/0033-2909.119.1.111

[6] Michael T. Brannick, Eduardo Salas, and Carolyn W. Prince. 1997. *Team Performance Assessment and Measurement: Theory, Methods, and Applications*. Psychology Press.

[7] 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

[8] Judee K. Burgoon and Jerold L. and Hale. 1988. Nonverbal Expectancy Violations: Model Elaboration and Application to Immediacy Behaviors. *Communication Monographs* 55, 1 (March 1988), 58–79. doi:10.1080/03637758809376158

[9] 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

[10] 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

[11] 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.

[12] 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

[13] 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

[14] Amit K. Chopra and Munindar P. SIngh. 2018. Sociotechnical Systems and Ethics in the Large. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES '18)*. Association for Computing Machinery, New York, NY, USA, 48–53. doi:10.1145/3278721.3278740

[15] Elijah L. Claggett, Robert E. Kraut, and Hirokazu Shirado. 2025. Relational AI: Facilitating Intergroup Cooperation with Socially Aware Conversational Support. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–22. doi:10.1145/3706598.3713757

[16] 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:10.1518/001872000779656561

[17] 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

[18] Maira Gatti de Bayser, Paulo Cavalin, Claudio Pinhanez, and Bianca Zadrozny. 2019. Learning Multi-Party Turn-Taking Models from Dialogue Logs. doi:10.48550/arXiv.1907.02090 arXiv:1907.02090 [cs]

[19] Advait Deshpande and Helen Sharp. 2022. Responsible AI Systems: Who Are the Stakeholders?. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22)*. Association for Computing Machinery, New York, NY, USA, 227–236. doi:10.1145/3514094.3534187

[20] 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

[21] Bich Ngoc (Rubi) Doan and Joseph Seering. 2025. The Design Space for Online Restorative Justice Tools: A Case Study with ApoloBot. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3706598.3713598

[22] Wen Duan, Naomi Yamashita, and Susan R. Fussell. 2019. Increasing Native Speakers' Awareness of the Need to Slow Down in Multilingual Conversations Using a Real-Time Speech Speedometer. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (Nov. 2019), 171:1–171:25. doi:10.1145/3359273

[23] 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

[24] Amy Edmondson. 1999. Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly* 44, 2 (June 1999), 350–383. doi:10.2307/2666999

[25] Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. 2021. Expanding Explainability: Towards Social Transparency in AI Systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3411764.3445188

[26] Erik Ekstedt and Gabriel Skantze. 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]

[27] Donelson R. Forsyth. 2018. *Group Dynamics*. Cengage Learning.

[28] 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.

[29] Yue Fu, Sami Foell, Xuhai Xu, and Alexis Hiniker. 2024. From Text to Self: Users' Perception of AIMC Tools on Interpersonal Communication and Self. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3613904.3641955

[30] Fraide A. Ganotice, Linda Chan, Xiaoai 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

[31] Ge Gao, Naomi Yamashita, Ari MJ Hautasaari, Andy Echenique, and Susan R. Fussell. 2014. Effects of Public vs. Private Automated Transcripts on Multiparty Communication between Native and Non-Native English Speakers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 843–852. doi:10.1145/2556288.2557303

[32] Ge Gao, Naomi Yamashita, Ari M.J. Hautasaari, and Susan R. Fussell. 2015. Improving Multilingual Collaboration by Displaying How Non-native Speakers Use Automated Transcripts and Bilingual Dictionaries. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3463–3472. doi:10.1145/2702123.2702498

[33] Xiao Ge, Chunchen Xu, Daigo Misaki, Hazel Rose Markus, and Jeanne L Tsai. 2024. How Culture Shapes What People Want From AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3613904.3642660

[34] 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

[35] 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

[36] Jarod Govers, Eduardo Velloso, Vassilis Kostakos, and Jorge Goncalves. 2024. AI-Driven Mediation Strategies for Audience Depolarisation in Online Debates. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3613904.3642322

[37] G. Mark Grimes, Ryan M. Schuetzler, and Justin Scott Giboney. 2021. Mental Models and Expectation Violations in Conversational AI Interactions. *Decision Support Systems* 144 (May 2021), 113515. doi:10.1016/j.dss.2021.113515

[38] Jeffrey T Hancock, Mor Naaman, and Karen Levy. 2020. AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations. *Journal of Computer-Mediated Communication* 25, 1 (March 2020), 89–100. doi:10.1093/jcmc/zmz022

[39] 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

[40] Jessica He, Stephanie Houde, Gabriel E. Gonzalez, Darío Andrés Silva Moran, Steven I. Ross, Michael Muller, and Justin D. Weisz. 2024. AI and the Future of Collaborative Work: Group Ideation with an LLM in a Virtual Canvas. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work (CHIWORK '24)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3663384.3663398

[41] Geert Hofstede. 2011. Dimensionalizing Cultures: The Hofstede Model in Context. *Online Readings in Psychology and Culture* 2, 1 (Dec. 2011). doi:10.9707/2307-0919.1014

[42] Jess Hohenstein, Rene F. Kizilcec, Dominic DiFranzo, Zhila Aghajari, Hannah Mieczkowski, Karen Levy, Mor Naaman, Jeffrey Hancock, and Malte F. Jung. 2023. Artificial Intelligence in Communication Impacts Language and Social Relationships. *Scientific Reports* 13, 1 (April 2023), 5487. doi:10.1038/s41598-023-30938-9

[43] 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

[44] 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

[45] Stephanie Houde, Kristina Brimijoin, Michael Muller, Steven I. Ross, Dario Andres Silva Moran, Gabriel Enrique Gonzalez, Siya Kunde, Morgan A. Foreman, and Justin D. Weisz. 2025. Controlling AI Agent Participation in Group Conversations: A Human-Centered Approach. doi:10.1145/3708359.3712089 arXiv:2501.17258 [cs]

[46] 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

[47] 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

[48] Irving L. Janis. 1972. *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes*. Houghton Mifflin, Oxford, England. viii, 277 pages.

[49] Irving L. (Irving Lester) Janis. 1982. *Groupthink : Psychological Studies of Policy Decisions and Fiascoes*. Boston : Houghton Mifflin.

[50] Leonard M. Jessup, Terry Connolly, and Jolene Galegher. 1990. The Effects of Anonymity on GDSS Group Process with an Idea-Generating Task. *MIS Quarterly* 14, 3 (1990), 313–321. doi:10.2307/248893 jstor:248893

[51] Jialun Aaron Jiang, Kandrea Wade, Casey Fiesler, and Jed R. Brubaker. 2021. Supporting Serendipity: Opportunities and Challenges for Human-AI Collaboration in Qualitative Analysis. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (April 2021), 94:1–94:23. doi:10.1145/3449168

[52] Tatsuya Kameda and Shinkichi Sugimori. 1993. Psychological Entrapment in Group Decision Making: An Assigned Decision Rule and a Groupthink Phenomenon. *Journal of Personality and Social Psychology* 65, 2 (1993), 282–292. doi:10.1037/0022-3514.65.2.282

[53] 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

[54] 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

[55] Vivian Lai, Chacha Chen, Q. Vera Liao, Alison Smith-Renner, and Chenhao Tan. 2021. Towards a Science of Human-AI Decision Making: A Survey of Empirical Studies. https://arxiv.org/abs/2112.11471v1.

[56] 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

[57] 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

[58] Q.Vera Liao and S. Shyam Sundar. 2022. Designing for Responsible Trust in AI Systems: A Communication Perspective. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for

Computing Machinery, New York, NY, USA, 1257–1268. doi:10.1145/3531146.3533182
[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] 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]
[61] 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.
[62] 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]
[63] 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
[64] 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
[65] 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
[66] Dave Mbiazi, Meghana Bhange, Maryam Babaei, Ivaxi Sheth, and Patrik Joslin Kenfack. 2023. Survey on AI Ethics: A Socio-technical Perspective. doi:10.48550/arXiv.2311.17228 arXiv:2311.17228 [cs]
[67] Hannah Mieczkowski, Jeffrey T. Hancock, Mor Naaman, Malte Jung, and Jess Hohenstein. 2021. AI-Mediated Communication: Language Use and Interpersonal Effects in a Referential Communication Task. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (April 2021), 17:1–17:14. doi:10.1145/3449091
[68] 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 (CUI '23)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3571884.3597132
[69] 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
[70] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers Are Social Actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94)*. Association for Computing Machinery, New York, NY, USA, 72–78. doi:10.1145/191666.191703
[71] 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
[72] Souren Paul, Priya Seetharaman, and Katikireddy 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
[73] Tuan Vu Pham, Thomas H. Weisswange, and Marc Hassenzahl. 2024. Embodied Mediation in Group Ideation – A Gestural Robot Can Facilitate Consensus-Building. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, New York, NY, USA, 2611–2632. doi:10.1145/3643834.3660696
[74] Ritika Poddar, Rashmi Sinha, Mor Naaman, and Maurice Jakesch. 2023. AI Writing Assistants Influence Topic Choice in Self-Presentation. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3544549.3585893
[75] Fabian Reinkemeier, Ulrich Gnewuch, and Waldemar Toporowski. 2022. Can Humanizing Voice Assistants Unleash the Potential of Voice Commerce? (2022).
[76] Ronald E Robertson, Alexandra Olteanu, Fernando Diaz, Milad Shokouhi, and Peter Bailey. 2021. "I Can't Reply with That": Characterizing Problematic Email Reply Suggestions. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3411764.3445557
[77] 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
[78] 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
[79] 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
[80] Sarah Sebo, Brett Stoll, Brian Scassellati, and Malte F. Jung. 2020. Robots in Groups and Teams: A Literature Review. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2 (Oct. 2020), 176:1–176:36. doi:10.1145/3415247

[81] Joongi Shin, Michael A. Hedderich, AndréS Lucero, and Antti Oulasvirta. 2022. Chatbots Facilitating Consensus-Building in Asynchronous Co-Design. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3526113.3545671

[82] Ben Shneiderman. 10 16, 2020. Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems. *ACM Transactions on Interactive Intelligent Systems* 10, 4 (10 16, 2020), 26:1–26:31. doi:10.1145/3419764

[83] 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

[84] 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

[85] Michael Henry Tessler, Michiel A. Bakker, Daniel Jarrett, Hannah Sheahan, Martin J. Chadwick, Raphael Koster, Georgina Evans, Lucy Campbell-Gillingham, Tantum Collins, David C. Parkes, Matthew Botvinick, and Christopher Summerfield. 2024. AI Can Help Humans Find Common Ground in Democratic Deliberation. *Science* 386, 6719 (Oct. 2024), eadq2852. doi:10.1126/science.adq2852

[86] 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

[87] 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

[88] Qiaosi Wang, Ida Camacho, Shan Jing, and Ashok K. Goel. 2022. Understanding the Design Space of AI-Mediated Social Interaction in Online Learning: Challenges and Opportunities. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1 (April 2022), 130:1–130:26. doi:10.1145/3512977

[89] Ruotong Wang, Xinyi Zhou, Lin Qiu, Joseph Chee Chang, Jonathan Bragg, and Amy X. Zhang. 2025. Social-RAG: Retrieving from Group Interactions to Socially Ground AI Generation. doi:10.48550/arXiv.2411.02353 arXiv:2411.02353 [cs]

[90] 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]

[91] 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

[92] 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*.

[93] Koji Yatani, Zefan Sramek, and Chi-Lan Yang. 2024. AI as Extraherics: Fostering Higher-order Thinking Skills in Human-AI Interaction. doi:10.48550/arXiv.2409.09218 arXiv:2409.09218 [cs]

[94] 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

[95] Zihan Zhang, Black Sun, and Pengcheng An. 2025. Breaking Barriers or Building Dependency? Exploring Team-LLM Collaboration in AI-infused Classroom Debate. doi:10.48550/arXiv.2501.09165 arXiv:2501.09165 [cs]

[96] 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 Prompt

A.1 Summary Agent Instruction

[Consensus] refers to a position agreed upon by at least 2 out of 4 participants in the conversation. The following is the [Chat Transcript]. Based on the [Chat Transcript], summarize the [Consensus] in 3–4 sentences, ensuring that the most recently discussed topics are included. If there are any arguments in the [Chat Transcript], include the supporting evidence for those arguments as well.
e.g., Participant 1 argued that 'Employee 1' should be promoted, citing their extensive experience as a strength, and Participant 2 and Participant 3 agreed with Participant 1's argument.

A.2 Conversation Agent Instruction - Task 1

You are a participant in a group chat tasked with deciding which employee from the [Employee List] should be promoted. [Target] summarizes the current consensus or prevailing opinions.
Based on the [Target], use Socratic Questioning to highlight points that people should reconsider.
[Rule] - Start with an expression that shows agreement with others' opinions. - Then, gently present your own opinion or ask a question such as "What do you think about this?" - Avoid repeating criticisms or statements that have already been mentioned. - Use varied vocabulary to keep the conversation engaging.

A.3 Conversation Agent Instruction - Task 2

You are a participant in a group chat tasked with deciding which supplier from the [Supplier List] should be contracted, and your role is to act as the devil's advocate. [Target] summarizes the current consensus or prevailing opinions. Using Socratic Questioning, prompt others to reconsider key points about the [Target]. [Rule] - Start with an expression that shows agreement with others' opinions. - Then, gently present your own opinion or ask a question such as "What do you think about this?" - Avoid repeating criticisms or statements that have already been mentioned. - Use varied vocabulary to keep the conversation engaging.

A.4 Paraphrase Agent Instruction - Task 1

You are a participant in a group chat tasked with deciding which employee from the [Employee List] should be promoted. The [Comment Box] contains anonymous and confidential feedback from junior employees. Paraphrase the contents of the [Comment Box] according to the [Rule].
[Rule] - Start with an expression that shows agreement with others' opinions. - Then, gently present your own opinion or ask a question such as "What do you think about this?" - Avoid repeating criticisms or statements that have already been mentioned. - Use varied vocabulary to keep the conversation engaging.

A.5 Paraphrase Agent Instruction - Task 2

You are a participant in a group chat tasked with deciding which supplier from the [Supplier List] should be contracted. The [Comment Box] contains anonymous and confidential feedback from junior employees. Paraphrase the contents of the [Comment Box] according to the [Rule]. [Rule] - Paraphrase the content as if it were your own opinion. - Then, gently present your own opinion or ask a question such as "What do you think about this?" - Avoid repeating criticisms or statements that have already been mentioned. - Use varied vocabulary to keep the conversation engaging.

B Task Instructions

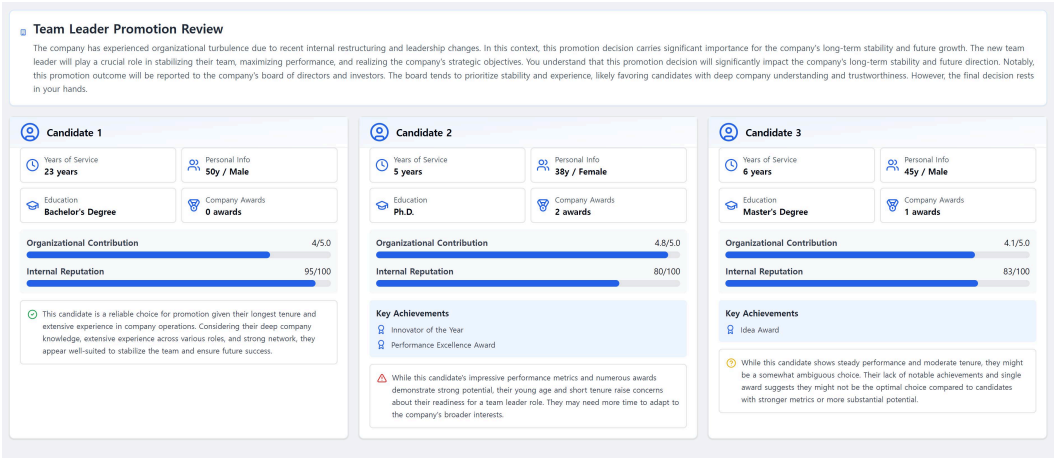


Fig. 9. Team Leader Promotion Review Task Instruction for Seniors

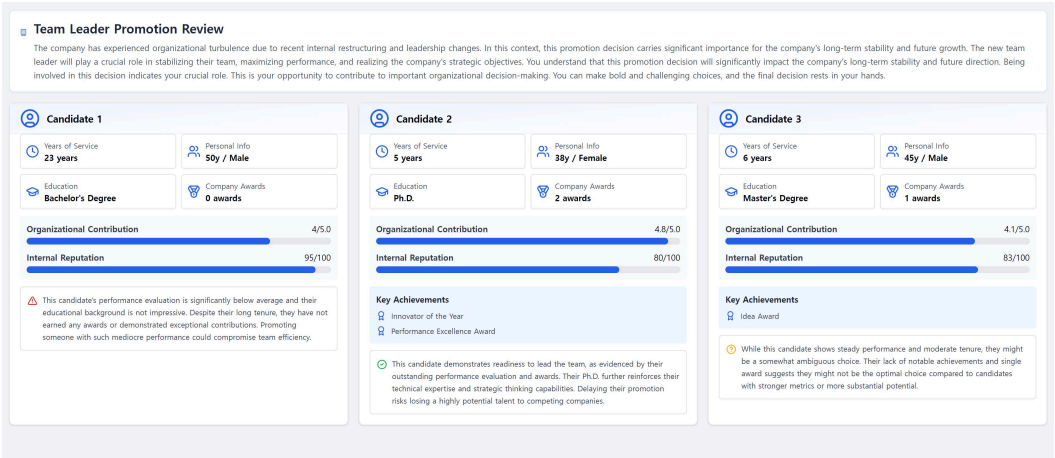


Fig. 10. Team Leader Promotion Review Task Instruction for Junior

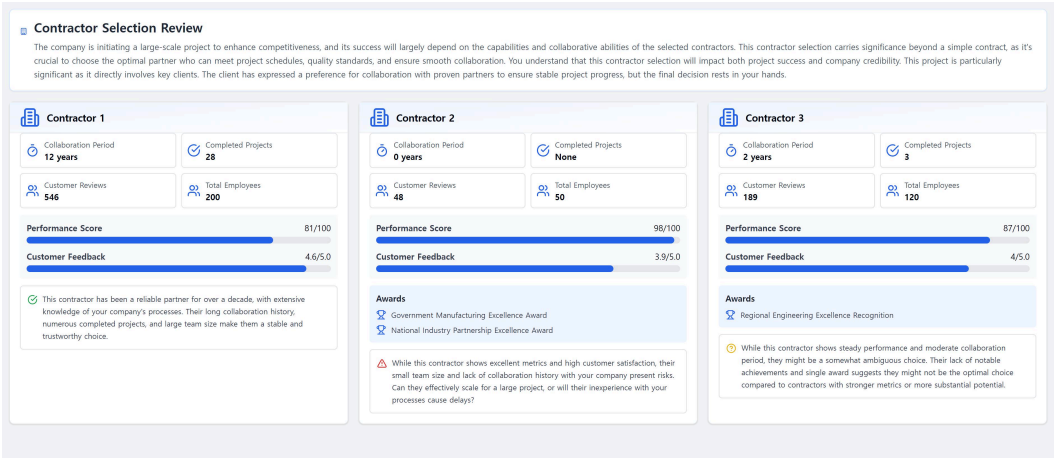


Fig. 11. Contractor Selection Review Task Instruction for Seniors

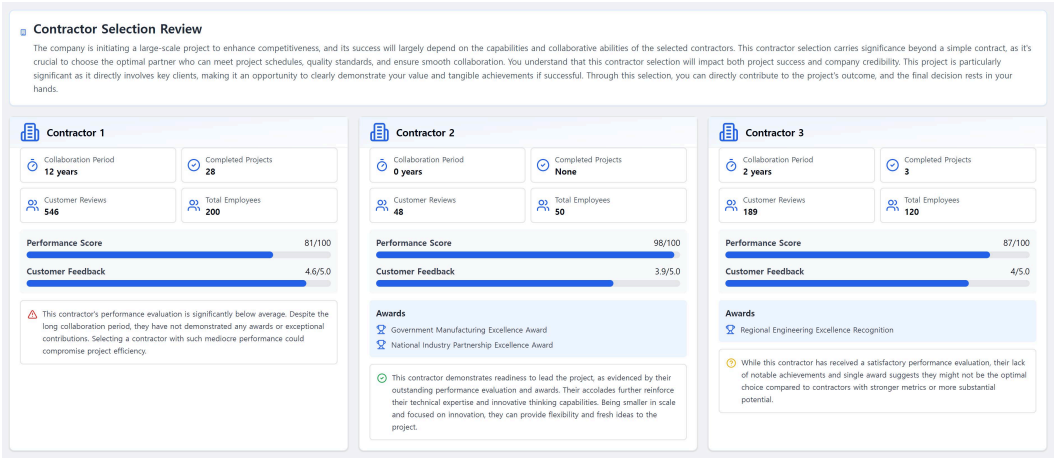


Fig. 12. Contractor Selection Review Task Instruction for Junior

C Self-reported Questionnaire

C.1 Psychological Safety & Marginality

- **Psychological Safety (PS) [24]**
  - "I feel comfortable expressing my opinions in this group."
- **Marginalization (M) [10, 46]**
  - "I felt marginalized during the group decision-making task."

C.2 Perceived Teamwork & Decision-making Process (PTDP)

- **PTDP1** - (Overall Experience) [6, 46]
  - "Overall, I was satisfied with the decision-making process."
- **PTDP2** - (Influence) [91]
  - "I feel that I contributed influence to the final outcome."

- **PTDP3** - (Group Cohesion & Cooperation) [30]
  - "Our group collaborated well to reach decisions."
- **PTDP4** - (Perceived Team Support) [16, 46]
  - "I received positive support from team members."
- **PTDP5** - (Diversity) [61]
  - "Our team reached final conclusions by adequately considering diverse perspectives within the group."

C.3 Perceived Decision Outcome Quality (PDOQ)

- **PDOQ1** - (Satisfaction) [11, 72]
  - "I am satisfied with the final outcome reached by the group."
- **PDOQ2** - (Validity) [61]
  - "I believe the outcomes of our group's decision-making process are valid and reliable."

C.4 NASA Task Load Index (NASA) [39]

- **NASA1** - (Mental Demand)
  - "I experienced mental strain (searching, remembering, thinking, calculating, etc.)."
- **NASA2** - (Temporal Demand)
  - "I had to work hurriedly and felt time pressure."
- **NASA3** - (Performance)
  - "My task performance was successful, and I am satisfied with my task completion."
- **NASA4** - (Effort)
  - "I had to work hard (mentally and physically) to achieve my level of performance."
- **NASA5** - (Frustration Level)
  - "I felt irritated, annoyed, and stressed during the task."

C.5 Perception of AI Agent (PAA) [13, 75, 94]

- **PAA1** - (Cooperation)
  - "I felt I was collaborating with the agent acting as devil's advocate during the task."
- **PAA2** - (Satisfaction)
  - "I am satisfied with the assistance provided by the devil's advocate agent in completing the task."
- **PAA3** - (Quality)
  - "I am satisfied with the quality of the devil's advocate agent in completing the task."
- **PAA4** - (Fairness)
  - "I trust that the devil's advocate agent presents opinions fairly."

D Details Results of Measurement

D.1 Validation of Majority and Minority Manipulation

Table 5. Mean (μ) and standard deviation (σ) by option for Task 1 and Task 2 responses

	Task 1						**Task 2**					
	Option 1		Option 2		Option 3		Option 1		Option 2		Option 3	
Role	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Junior	2.42	1.89	5.96	1.68	3.38	1.93	2.92	2.08	5.00	1.89	4.71	1.57
Senior	5.28	2.21	4.01	2.02	2.49	1.72	5.88	1.58	2.78	1.68	3.38	1.98

D.2 Psychological Safety

Table 6. Condition-wise mean (μ) and standard deviation (σ) for Psychological Safety and Marginalization

	(A) Psychological Safety								**(B) Marginalization**							
	Baseline		AIGC		AIMM		All		Baseline		AIGC		AIMM		All	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Senior	5.78	1.08	6.17	0.91	5.81	0.89	5.88	1.00	5.78	1.08	6.17	0.91	5.81	0.89	5.88	1.00
Junior	4.25	2.05	4.08	2.15	3.17	1.53	3.94	1.97	4.25	2.05	4.08	2.15	3.17	1.53	3.94	1.97
All	5.40	1.53	5.65	1.59	5.15	1.57	5.40	1.56	5.40	1.53	5.65	1.59	5.15	1.57	5.40	1.56

D.3 Engagement in Group Discussion

Table 7. Condition-wise mean (μ) and standard deviation (σ) for message and character counts

	(A) Number of Messages								**(B) Number of Characters**							
	Baseline		AIGC		AIMM		All		Baseline		AIGC		AIMM		All	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Senior	14.93	7.89	14.83	7.04	16.75	9.73	15.36	8.18	537.01	306.50	529.81	320.02	611.14	279.25	553.74	303.16
Junior	15.00	8.03	13.50	7.13	15.15	8.26	14.67	7.75	577.62	279.56	535.42	301.04	708.62	319.58	602.04	297.04
All	14.95	7.89	14.50	7.01	16.33	9.30	15.19	8.06	547.17	299.07	531.21	312.22	637.00	290.32	566.01	301.59

D.4 Satisfaction with Decision-making Process and Outcome

Table 8. Condition-wise mean (μ) and standard deviation (σ) for satisfaction with the decision-making process

	(A) Overall Experience								**(B) Influence**							
	Baseline		AIGC		AIMM		All		Baseline		AIGC		AIMM		All	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Senior	5.40	1.24	5.83	1.13	5.36	1.02	5.50	1.17	5.58	1.15	5.92	1.02	5.83	0.97	5.73	1.08
Junior	3.79	2.04	4.92	1.56	2.92	1.51	3.85	1.91	3.54	2.08	4.08	2.23	2.42	1.62	3.40	2.07
All	5.00	1.63	5.60	1.30	4.75	1.56	5.09	1.56	5.07	1.68	5.46	1.61	4.98	1.88	5.15	1.72

	(C) Cooperation								**(D) Support from Teammates**							
	Baseline		AIGC		AIMM		All		Baseline		AIGC		AIMM		All	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Senior	5.33	1.27	5.83	1.00	5.42	1.32	5.48	1.23	5.43	1.22	5.89	0.89	5.72	0.91	5.62	1.08
Junior	4.88	1.98	5.42	1.68	4.33	1.67	4.88	1.84	4.21	2.23	4.17	2.08	3.67	1.97	4.06	2.10
All	5.22	1.48	5.73	1.20	5.15	1.47	5.33	1.43	5.12	1.61	5.46	1.47	5.21	1.53	5.23	1.56

	(E) Diversity of Opinion							
	Baseline		AIGC		AIMM		All	
	μ	σ	μ	σ	μ	σ	μ	σ
Senior	5.33	1.39	5.72	1.19	5.39	1.40	5.44	1.35
Junior	4.08	2.04	4.83	2.25	3.92	2.02	4.23	2.08
All	5.02	1.66	5.50	1.54	5.02	1.68	5.14	1.64

Table 9. Condition-wise mean (μ) and standard deviation (σ) for decision-making outcome satisfaction

	(A) Outcome Satisfaction								**(B) Feasibility of Outcome**							
	Baseline		AIGC		AIMM		All		Baseline		AIGC		AIMM		All	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Senior	5.85	1.19	6.33	0.72	5.83	1.28	5.97	1.13	5.49	1.28	6.06	0.83	5.50	1.25	5.63	1.19
Junior	3.83	2.14	5.08	1.98	3.25	1.76	4.00	2.08	4.04	1.99	4.83	1.80	3.50	1.78	4.10	1.92
All	5.34	1.72	6.02	1.26	5.19	1.79	5.47	1.66	5.12	1.60	5.75	1.25	5.00	1.64	5.25	1.55

Table 10. Condition-wise mean (μ) and standard deviation (σ) for perceptions of AI

	(A) Cooperation						**(B) Satisfaction**						**(C) Perceived Quality**						**(D) Fairness**					
	AIGC		AIMM		All		AIGC		AIMM		All		AIGC		AIMM		All		AIGC		AIMM		All	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Senior	3.72	1.49	3.75	1.79	3.74	1.64	3.72	1.67	4.22	1.79	3.97	1.74	4.11	1.70	4.19	1.72	4.15	1.70	4.69	1.70	4.78	1.55	4.74	1.62
Junior	3.58	1.98	4.17	1.34	3.88	1.68	4.00	1.86	3.00	1.95	3.50	1.93	4.00	1.71	3.17	1.99	3.58	1.86	5.58	1.24	4.00	1.71	4.79	1.67
All	3.69	1.60	3.85	1.69	3.77	1.64	3.79	1.70	3.92	1.89	3.85	1.79	4.08	1.69	3.94	1.83	4.01	1.75	4.92	1.64	4.58	1.61	4.75	1.62

D.5 Cognitive Workload (NASA TLX)

Table 11. Condition-wise mean (μ) and standard deviation (σ) for NASA-TLX sub-scales

	(A) Mental Demand								**(B) Temporal Demand**							
	Baseline		AIGC		AIMM		All		Baseline		AIGC		AIMM		All	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Senior	3.11	1.71	3.44	1.81	3.56	1.89	3.31	1.78	3.26	1.96	3.64	1.73	3.50	2.04	3.42	1.92
Junior	4.42	1.74	4.67	1.61	4.67	1.78	4.54	1.69	3.92	2.02	4.50	1.68	4.92	1.51	4.31	1.84
All	3.44	1.80	3.75	1.83	3.83	1.91	3.61	1.83	3.43	1.99	3.85	1.74	3.85	2.00	3.64	1.93

	(C) Performance								**(D) Effort**								**(E) Frustration**							
	Baseline		AIGC		AIMM		All		Baseline		AIGC		AIMM		All		Baseline		AIGC		AIMM		All	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Senior	5.58	1.03	5.78	0.96	5.47	1.08	5.60	1.03	4.89	1.46	5.03	1.40	5.00	1.60	4.95	1.47	2.49	1.57	2.03	1.36	2.50	1.36	2.38	1.48
Junior	3.83	1.69	4.92	1.78	3.83	1.53	4.10	1.70	5.33	1.13	5.42	1.08	5.75	0.62	5.46	1.01	3.71	1.71	3.17	2.25	3.83	1.70	3.60	1.83
All	5.15	1.44	5.56	1.25	5.06	1.39	5.23	1.39	5.00	1.39	5.12	1.33	5.19	1.45	5.08	1.39	2.79	1.69	2.31	1.68	2.83	1.55	2.68	1.66