"Shall We Dig Deeper?": Designing and Evaluating Strategies for LLM Agents to Advance Knowledge Co-Construction in Asynchronous Online Discussions

Yuanhao Zhang
CSE
Hong Kong University of Science and Technology
Hong Kong, China
yzhangiy@connect.ust.hk

Wenbo Li
Department of Computer Science
North Carolina State University
Raleigh, North Carolina, USA
wli55@ncsu.edu

Xiaoyu Wang
School of Engineering
The Hong Kong University of Science and Technology
Hong Kong, Hong Kong
xwangij@connect.ust.hk

Kangyu Yuan
Computer Science and Engineering
Hong Kong University of Science and Technology
Hong Kong, China
kyuanaf@connect.ust.hk

Shuai Ma
Institute of Software, Chinese Academy of Sciences
Beijing, China
mashuai@iscas.ac.cn

Xiaojuan Ma
Hong Kong University of Science and Technology
Hong Kong, Hong Kong
mxj@cse.ust.hk

Abstract

Asynchronous online discussions enable diverse participants to co-construct knowledge beyond individual contributions. This process ideally evolves through sequential phases, from superficial information exchange to deeper synthesis. However, many discussions stagnate in the early stages. Existing AI interventions typically target isolated phases, lacking mechanisms to progressively advance knowledge co-construction, and the impacts of different intervention styles in this context remain unclear and warrant investigation. To address these gaps, we conducted a design workshop to explore AI intervention strategies (task-oriented and/or relationship-oriented) throughout the knowledge co-construction process, and implemented them in an LLM-powered agent capable of facilitating progression while consolidating foundations at each phase. A within-subject study (N=60) involving five consecutive asynchronous discussions showed that the agent consistently promoted deeper knowledge progression, with different styles exerting distinct effects on both content and experience. These findings provide actionable guidance for designing adaptive AI agents that sustain more constructive online discussions.

CCS Concepts

• **Human-centered computing → Interactive systems and tools**; *Empirical studies in HCI*.

Keywords

Knowledge Co-Construction, Human-AI Collaboration, Conversational Agent

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3790551

ACM Reference Format:
Yuanhao Zhang, Wenbo Li, Xiaoyu Wang, Kangyu Yuan, Shuai Ma, and Xiaojuan Ma. 2026. "Shall We Dig Deeper?": Designing and Evaluating Strategies for LLM Agents to Advance Knowledge Co-Construction in Asynchronous Online Discussions. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 27 pages. https://doi.org/10.1145/3772318.3790551

1 Introduction

Asynchronous online discussion platforms (e.g., Reddit [109], Discourse [34]) have become essential venues for knowledge co-construction [39, 134, 137], a process that enables participants from diverse backgrounds to collaboratively examine, extend, and refine one another's ideas, thereby enhancing shared understanding of complex issues across various disciplines [72, 89, 141] (Figure 1). Compared with individually held knowledge, such collective processes integrate multiple perspectives, reconcile divergent viewpoints, and generate more robust and comprehensive understandings of a subject [39, 50, 56]. Ideally, knowledge co-construction unfolds through a staged interactive progression, evolving from initial idea sharing (*initiation*), building on and enriching each other's contributions (*exploration*), explicitly addressing divergent perspectives to reach shared understanding (*negotiation*), to finally synthesizing and reflecting jointly created knowledge (*co-construction*) [51, 104, 117]. While surface-level exchanges suffice for socializing or simple information sharing, it is critical for discussions centered on complex or open-ended topics to achieve the necessary depth [28]. However, prior empirical research indicated that many such asynchronous discussions remain concentrated in the early, surface-level stages, with limited progression toward more integrative phases [39, 89, 102]. For users, shallow exchanges deprive contributors of meaningful uptake – leaving their inputs unexamined, unelaborated, and unintegrated – while consumers struggle to obtain coherent explanations or reconciled viewpoints, instead encountering fragmented comments [13, 27, 82]. For platforms, such

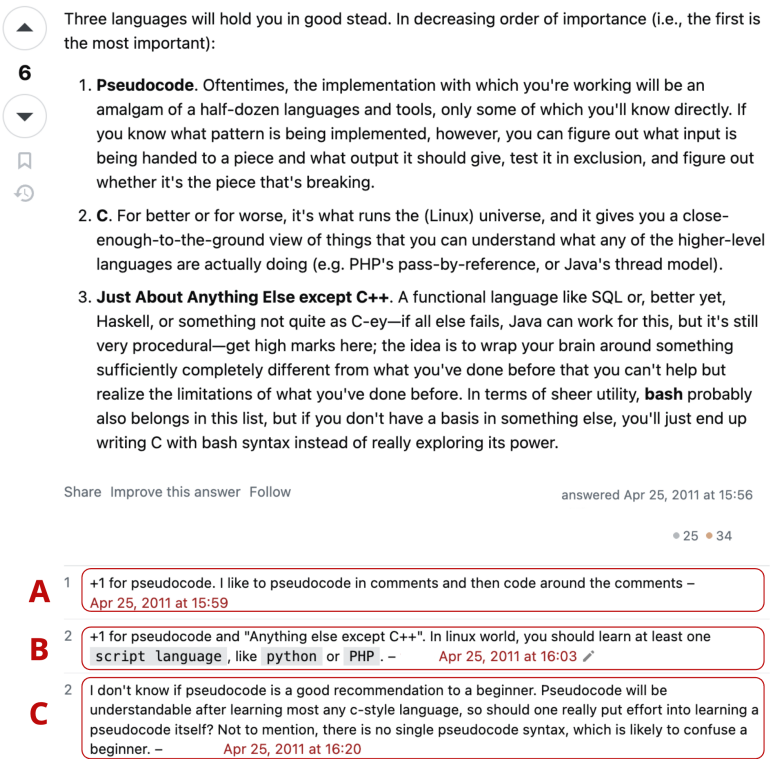


Figure 1: The screenshot shows a Stack Exchange discussion thread responding to the question "What language for starting on Linux?", where the accepted answer (top) recommends learning pseudocode, C, and any scripting language except C++. Comments A, B, and C co-construct knowledge by complementing, elaborating, or challenging this suggestion. Comment A agrees with and expands on the value of pseudocode. Comment B reinforces the recommendation while adding specific language examples such as Python and PHP. In contrast, Comment C challenges the practicality of learning pseudocode for beginners and raises concerns about its lack of standard syntax. Together, these interactions illustrate how community members collaboratively refine and contest knowledge claims, contributing to a richer collective understanding.

stagnation undermines their value as sustainable knowledge repositories, producing piecemeal discussions rather than synthesized collective knowledge [27]. Therefore, the need arises for dynamic support for users of these platforms to achieve sustained and meaningful collective knowledge building as their discussions unfold [84].

Advances in AI technology have spurred the deployment of agents beyond one-to-one interactions to multi-party discussions, where they employ various strategies to support and shape group dynamics [35, 122]. One type of agents aims at interactional coordination and behavior management, such as ensuring balanced participation [8, 36, 75], emotion regulation [12, 107, 130], and discussion pacing [76, 113]. They helped smooth the discussion process, but paid limited attention to the depth and breadth of content development. Another line of automated interventions focuses directly on supporting knowledge co-construction, for instance by persistently challenging claims [24, 90] or prompting clarifications [21, 52, 68]. These interventions typically focus on content issues within a particular phase, without a broader, process-orchestrated emphasis on fostering the progression across phases. Thus, the design of strategies for meaningfully advancing knowledge co-construction remains largely underexplored.

Previous studies suggest that humans exercise different styles of interventions to promote knowledge co-construction, generally distinguished between task-oriented approaches (providing guidance to advance knowledge-building [101, 129]) and relationship-oriented approaches (contributing alongside others to foster closer and more equal collaboration [24, 140]). These facilitation styles are likely to shape group dynamics, motivation, and knowledge-sharing outcomes in distinct ways [32, 108]. While these findings concern human facilitation, prior studies indicate that agents engaging in multi-party interactions with different intervention styles can elicit varied reactions from human participants [74, 136]. These insights motivate our investigation into how different AI intervention styles may affect the knowledge co-construction dynamics around content development and participant experiences during asynchronous online discussions.

To this end, we first conducted a design workshop with 12 participants (nine active contributors to online knowledge discussions and three AI designers) to co-design the intervention strategies an AI member can take to advance knowledge development across

different co-construction phases. The strategies were designed in telling, selling, participating, and delegating styles, respectively. Participants first identified the sufficiency criteria for each phase, then brainstormed and calibrated strategies that an AI agent with a given style could employ to achieve these criteria. After refinement and synthesis of these insights, we further developed an LLM-powered agent to progressively advance online knowledge co-construction, ensuring robust foundations at each phase. The agent can analyze each comment's characteristics within the evolving discussion, monitor phase-level dynamics, and operationalize workshop-derived, phase-specific interventions in different styles. With the agent, we proceeded to addressing the following research questions:

RQ1 How do the AI intervention styles influence the progression of knowledge co-construction in asynchronous online discussions?

RQ2 How do the intervention styles shape human perceptions of the AI agent and their user experience?

RQ3 How do the intervention styles affect human-human interaction in asynchronous online discussions?

To answer these questions, we conducted a within-subject study (N = 60) where each participant engaged in five consecutive discussion threads, each involving six participants, under five conditions (four AI intervention styles and one human-only baseline). Mixed-methods analyses – including thread-level metrics, in-task surveys, and interviews – revealed that compared with the baseline, agents employing telling, selling, and participating styles had the potential to advance the knowledge co-construction into deeper phases and to establish a stronger foundation in the *exploration* phase, with each style exhibiting distinct strengths and limitations in driving such advancements. These three styles were also perceived more favorably than the delegating style, with the participating style receiving the highest appreciation. Participants also reported that different styles influenced human–human interactions in distinct ways.

The contributions of our work are threefold. First, we design an actionable set of phase-tailored intervention strategies in different styles for AI agents to advance knowledge co-construction as active discussion participants. Second, our mixed-methods evaluation provides empirical insights into the impacts of various AI intervention styles on knowledge co-construction and participants' experiences. Third, we distill key design implications to guide the development of adaptive AI agents for multi-party interactions.

2 Related Work

2.1 Knowledge Co-Construction in Asynchronous Discussion Platforms

Asynchronous discussion forums (e.g., Reddit [109], Discourse [34], Stack Exchange [116]) have become crucial online platforms where millions of users with diverse experiences and expertise engage in knowledge seeking and information sharing [6, 38]. Unlike ephemeral real-time conversations, asynchronous forums transform user interactions into persistent knowledge repositories [26, 40]. Posts and replies, preserved as textual records, allow continuous access, quotation, and iterative elaboration, thus establishing a cumulative community knowledge base [40]. Additionally, asynchronous forums offer temporal flexibility, enabling participants to thoughtfully contribute at their own pace [3]. These characteristics make asynchronous forums particularly effective for harnessing the "wisdom of crowds" and supporting knowledge co-construction.

Knowledge co-construction refers to a collaborative process wherein individuals not only share information but also collectively refine, negotiate, and expand each other's ideas, fostering collective understanding beyond individual contributions [39, 77, 137]. It can be conceptualized as a progressive, multi-phase process, evolving from superficial information sharing toward deeper integration [51, 104, 117]. Among existing frameworks describing this process, the Interaction Analysis Model (IAM) [51] is widely recognized. IAM characterizes knowledge co-construction as five sequential phases: (1) sharing ideas, (2) exploring dissonance, (3) negotiating meaning, (4) testing tentative knowledge syntheses, and (5) applying newly constructed knowledge. Lucas et al. [89] empirically assessed IAM and suggested that higher phases may require reconsideration or consolidation. In line with this perspective, Onrubia et al. [104] proposed a simplified four-phase analytical model inspired by IAM, comprising phases of (1) initiation, (2) exploration, (3) negotiation, and (4) co-construction. Stahl's collaborative knowledge-building model [117] also shares conceptual similarities, emphasizing the progression from publicly stated ideas through argumentation development, negotiation toward shared understanding, and eventually the objectification of collaboratively generated knowledge. Building upon these empirically and theoretically validated models, this paper employs a refined four-phase coding scheme (detailed in Section 3.1) to analyze knowledge co-construction in our study.

Empirical studies analyzing asynchronous knowledge co-construction using these frameworks consistently observed that interactions predominantly occur at superficial levels, with significantly fewer interactions reaching the deeper phases [88, 89, 144]. For instance, Dubovi et al. [39], employing IAM to analyze comments on YouTube videos, reported that most interactions remained at initial sharing and exploration phases. McLoughlin et al. [96] found similar patterns in their analysis of university students' interactions in online learning forums. Such stagnation limits contributors' opportunities to have their ideas examined and built upon, discouraging sustained engagement [13, 82]; leaves consumers with fragmented and sometimes contradictory information that hinders coherent understanding [26, 39]; and reduces platforms' value as repositories of synthesized, durable knowledge [27]. To address this gap, our work explores the potential of AI agents to engage in multi-party interactions as active discussion participants, strategically fostering more advanced phases of knowledge co-construction.

2.2 Interventions to Support Knowledge Co-construction

Extensive research in the learning sciences demonstrates that human intervention is critical for fostering productive knowledge co-construction [73, 128], which generally fall into two categories. The first is instructional intervention, where facilitators actively guide the discussion with a strong task focus. For instance, teachers can prompt students to elaborate on their explanations [133] or ask

specific follow-up questions to deepen the discussion [132], thereby driving the group toward learning goals. The second adopts a more participatory approach, characterized by egalitarian, teammate-like relationships. In this mode, teachers can act as "role models" by demonstrating example behaviors [123, 128], or, as proposed by King [78], peers can assess and tutor one another to foster mutual knowledge construction.

With advances in Natural Language Processing (NLP), AI agents have increasingly been deployed to support multi-party interactions. A significant strand of this research focuses on agents acting as coordinators, emphasizing interaction management over substantive knowledge building. For instance, agents like IntroBot [113] and BlahblahBot [114] focus on social lubrication (e.g., icebreakers), while others like GroupfeedBot [75] regulate participation rates or manage group emotion [12, 107]. While effective for maintaining productive dynamics, these "coordinator" agents mainly manage the flow of conversation without enhancing the quality of the knowledge being constructed.

To directly support knowledge co-construction, more recent work has begun to assume the intervention roles traditionally held by humans. Mirroring the strategies above, these AI interventions generally adopt two distinct approaches. The first positions the AI as a facilitator that directs the knowledge-building process. For instance, some agents [21, 68, 101] post follow-up questions to prompt users to refine and reflect on their prior statements. Other agents steer groups toward deeper collaboration by delivering timely prompts (e.g., JIA [37]) or enriching discussions with context-relevant informational suggestions (e.g., PAPERPING [129]). The second category positions the AI as a teammate [140], which actively contribute task knowledge rather than providing guidance. For instance, Chiang et al. [24] developed a "devil's advocate" agent that provides arguments from a position opposing the majority opinion. However, these interventions typically address isolated conversational moments, without a process-orchestrated paradigm to adapt their support to the evolving stages of knowledge co-construction. This study addresses this gap by conducting a Design Workshop to develop phase-specific AI intervention strategies and a narrative literature review to derive empirically grounded design requirements for systematically advancing the depth and quality of collective knowledge in asynchronous discussions.

2.3 Frameworks of Intervention in Group Settings

Designing AI intervention strategies for multi-party interaction can be informed by models validated in human group contexts [23, 48]. Task-Relationship (TR) framework depicts two distinct behavioral dimensions commonly used to categorize human intervention styles [119]: **task** orientation, which emphasizes providing direction and achieving goals, and **relationship** orientation, which emphasizes fostering interpersonal support and active collaboration through shared responsibility. Together, these dimensions form a descriptive framework characterizing most observed intervention behaviors in group settings [45]. Several influential models further conceptualized specific styles within the space defined by these two dimensions through a quadrant logic (e.g., Managerial Grid Model[14], Situational Leadership Model (SLM) [42, 63]).

TR framework also effectively characterize the landscape of AI interventions for knowledge co-construction reviewed in Section 2.2. Specifically, facilitator-like agents [37, 101, 129], which focus on providing guidance and structuring the discussion to foster knowledge building, are more **task**-orientated. The teammate-like roles [24, 140], which build rapport and equal relationship by acting as "lead-by-example" peers (i.e., actively contributing knowledge alongside others for demonstration), are more **relationship**-oriented. Due to this conceptual alignment, this work contextually adapts the TR framework to define AI intervention styles. Drawing on the practices of prior models [42, 63], we also apply the quadrant logic-mapping the intersection of these two dimensions—to systematically explore how different combinations of guidance (task) and peer-like participation (relationship) influence knowledge co-construction.

Research indicates that intervention style can shape group outcomes. For example, Lu et al. [88] found that adopting a peer-like role—where teachers actively contribute to the discussion alongside students—led to higher-quality knowledge construction than merely facilitating. In contrast, Raza et al. [108] reported greater group achievement from more task-oriented interventions. While these findings concern human intervention, prior studies also suggest that participants are sensitive to the styles used by AI agents in group settings and may respond differently depending on the style [74, 136]. Even though AI agents act as equal members in asynchronous discussions, they become a temporary, ad hoc leader when trying to engage other members to advance knowledge co-construction. Like humans, AI agents have the choice to prioritize task completion (i.e., driving the group toward synthesizing collective knowledge) and/or relationship (i.e., nurturing mutual engagement and shared ownership of the knowledge), yet their impacts in knowledge building are underexplored. This gap motivates our investigation into how AI agents, when adopting these intervention styles, influence both the dynamics of knowledge co-construction and participants' experiences in asynchronous forums.

3 Design and Implementation

This section illustrates the design and implementation of AI intervention strategies. Based on theoretical frameworks, we conducted a design workshop to identify agent strategies for advancing knowledge co-construction across different styles. Insights from the workshop informed the development of an LLM-based agent.

3.1 Theoretical Framework

The design of strategies across different phases and styles requires a theoretical grounding in process models of knowledge co-construction and established frameworks of intervention styles. Drawing on IAM [51], Onrubia et al. [104], and Stahl's collaborative knowledge-building model [117], we adopt a refined four-phase coding scheme to examine knowledge co-construction in asynchronous discussions:

Phase1 **Initiation**: Introducing novel perspectives without engaging in interactive discourse with others.

Phase2 **Exploration**: Building upon or complementing prior contributions with their own information.

Phase3 Negotiation: Explicitly addressing the inconsistencies identified in earlier discussions.

Phase4 Co-construction: Collaboratively integrating negotiation insights into a coherent synthesis, and reflecting on the jointly created knowledge.

Note that while these phases model an ideal progression of knowledge construction, real-world discourse is often iterative rather than strictly linear. Discussions frequently cycle between phases—particularly exploration and negotiation—as new participants enter.

For intervention styles, we apply quadrant logic to the TR framework [119], mapping the intersection of **task** and **relationship** dimensions. We adapt the resulting four styles [1] to the context of knowledge co-construction:

- **Telling**: High-task, low-relationship – the agent provides explicit instructions to guide knowledge co-construction.
- **Selling**: High-task, high-relationship – the agent suggests directions for knowledge co-construction and actively persuades others to adopt them, fostering shared commitment to the emerging knowledge.
- **Participating**: Low-task, high-relationship – the agent reduces directive behavior, instead collaborating with the group and sharing responsibility for constructing knowledge.
- **Delegating**: Low-task, low-relationship style – the agent adopts a hands-off approach, granting autonomy to members while monitoring progress with minimal intervention.

3.2 Design Workshop

We conducted a design workshop to identify agent strategies for advancing knowledge co-construction. While the theoretical framework are grounded in prior literature, operationalizing these concepts requires capturing the situated practices of online communities. This participatory approach complements the literature by surfacing contributors' tacit knowledge, ensuring our design is ecologically valid and aligned with real-world dynamics.

3.2.1 Participants and Procedure. With IRB approval, we conducted a three-hour online design workshop with 12 participants (four female, eight male), recruited via online advertisements, social media, and word-of-mouth. The group comprised three AI designers—two of whom specialize in AI in Education with backgrounds in learning sciences—and nine frequent contributors to online knowledge discussions (three contributed daily, four multiple times a week, and two at least once a week; summarized in Table 1).

The workshop was held via a video-conferencing platform [1], with a shared online whiteboard tool [97] for collaborative editing. It began with a 15-minute briefing on the definitions of the four knowledge co-construction phases and the four intervention styles, supplemented by concrete examples to ensure a shared theoretical foundation. Participants then engaged in three sequential activities:

- **Activity 1: Defining Sufficiency Criteria for Each Phase (15 minutes)** Building on theoretical definitions, participants collaboratively specified the sufficiency criteria for each phase. They shared perspectives, discussed differences, and consolidated these into an agreed set of outcomes, which served as design targets for subsequent activities. The results were documented in the shared spreadsheet.
- **Activity 2: Brainstorming Phase-Specific Intervention Strategies (60 minutes)** Participants were divided into three groups (each comprising one AI designer and three frequent contributors), each assigned one intervention style—Telling, Selling, or Participating. In breakout rooms with a facilitator, groups brainstormed agent strategies that could support phase transitions and achieve the sufficiency criteria from Activity 1. Contributions could be abstract strategy descriptions or concrete message examples, informed by participants' own discussion experiences. Outputs were recorded in the shared document. We excluded the Delegating style since its low-task, low-relationship orientation inherently requires minimal agent intervention, allowing it to be implemented through simple, predefined phase-opening prompts without the need for extensive co-design.
- **Activity 3: Consolidating and Refining Representative Strategies (75 minutes)** All participants reconvened to exchange and refine ideas generated in Activity 2, guided by a facilitator. New suggestions were welcomed during this stage. The process resulted in a final set of representative intervention strategies for each style (except delegating) across all four phases.

The workshop was audio-recorded and transcribed, and all participant-generated documents were retained. Two researchers independently analyzed the transcripts and artifacts, iteratively discussing and refining both the sufficiency criteria (from Activity 1) and the strategies (from Activities 2–3) to ensure that they accurately represented participants' intended meanings and design rationales. Discrepancies were resolved through discussion until consensus was reached.

3.2.2 Summary of Findings. Table 2 summarizes the sufficiency criteria for each phase and the representative intervention strategies for each agent style. Given the collaborative structure of the workshop, results reflect group consensus rather than individual frequencies: sufficiency criteria were established by the full cohort ($N = 12$), while strategies were synthesized in subgroups ($N = 4$) and subsequently validated by all participants ($N = 12$). Below, we detail how these outcomes manifest across the four phases.

Initiation: The desired outcome of this phase is to cultivate a broad range of starting points, offering a foundation for subsequent interactions. To support this, telling-style agents were envisioned as issuing directives that introduce novel starting ideas not yet present in the conversation, prompting others to discuss them. Selling-style agents would propose such ideas while emphasizing their potential relevance or value. Participating-style agents would contribute novel perspectives of their own.

Exploration: This phase focuses on expanding existing contributions, ensuring that most are substantively addressed. Workshop participants suggested assessing sufficiency by examining the completeness of the argumentative structure. Telling-style agents could identify underexplored arguments and instruct participants to elaborate on them from specific missing angles. Selling-style agents would encourage such elaboration by underscoring its potential

[1] Our definitions are a contextual adaptation of the TR framework; similarity to SLM nomenclature does not imply an adopting its original definitions.

ID	Gender	Primary Platform	Freq. of Contribution	Confidence Level
PC1	Male	Zhihu	At least once a week	5
PC2	Male	Reddit	Multiple times a week	5
PC3	Female	Zhihu	Daily	6
PC4	Female	Quora	Multiple times a week	7
PC5	Female	Stack Exchange	Multiple times a week	6
PC6	Female	Stack Exchange	Multiple times a week	6
PC7	Male	Zhihu	At least once a week	4
PC8	Male	Quora	Daily	6
PC9	Male	Reddit	Daily	5

Table 1: Demographics and online behaviors of the active contributors in the workshop. Confidence level is self-reported on a 7-point Likert scale (1=Low, 7=High)

.

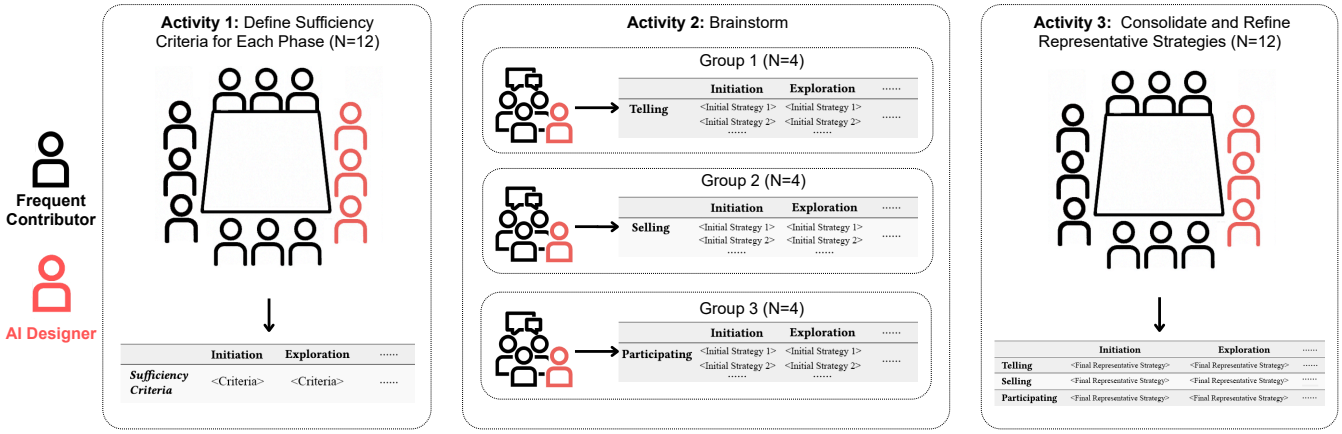


Figure 2: Activities of the workshop. Sufficiency criteria for each phase were defined by the full cohort (N = 12), while representative intervention strategies for each agent style were first synthesized in subgroups (N = 4) and then validated by all participants (N = 12)

significance. Participating-style agents extend these contributions themselves, contributing complementary content to foster further exploration.

Negotiation: In this phase, sufficiency is indicated when inconsistencies emerged earlier have been resolved, with consensus reached on each issue. However, achieving full agreement may not always be the ideal outcome, such as when it risks suppressing healthy disagreement or forcing premature convergence. Therefore, a valid "consensus" in this context also encompasses states where distinct viewpoints are clearly clarified and acknowledged. Participants noted that agents could identify unresolved disagreements and stimulate discussion towards such resolution. Telling-style agents could propose concrete approach for reaching consensus and ask participants to engage accordingly. Selling-style agents would pair this guidance with persuasion about the value of resolving the issue (e.g., benefits, risks of deferral). Participating-style agents present their own stance as a starting point for engaging others in co-developing negotiated agreements.

Co-Construction: The final phase focuses on collaboratively synthesizing agreements reached and linking them to practical applications or reflective insights. Participants identified the active process of synthesis—rather than just the resulting artifact—as the critical mechanism for solidifying consensus. Consequently, success was defined as a collective synthesis where contributors explicitly articulate the agreed knowledge, supported by substantial reflection. In this phase, telling-style agents explicitly instruct participants to produce a holistic summary, specifying which issues to address and from which perspectives to reflect. Selling-style agents, in addition to providing guidance, advocate for the value of comprehensive summarizing and reflection. Participating-style agents, acknowledging the potential labor burden, initiate a partial summary with their own reflections to scaffold the process, yet intentionally leave gaps for users to complete and enrich.

3.3 Agent Design and Implementation

To evaluate the derived strategies, we developed an LLM agent that operationalizes them in asynchronous discussion contexts. Drawing

	Initiation	Exploration	Negotiation	Co-Construction
Sufficiency Criteria	A varied set of initial contributions that invite further exploration.	Most existing contributions are substantively addressed, showing a well-developed argumentative structure.	All identified conflicts are resolved, with consensus on each issue.	A collective synthesis capturing all agreements and linking to reflective insights.
Telling	Direct others toward unexplored starting ideas for discussion.	Instruct participants to elaborate on underexplored arguments from missing angles.	Propose concrete angles on the conflict for reaching consensus and prompt participants to address them.	Direct participants to produce a complete summary and specify connections to potential applications or lessons learned.
Selling	Propose new directions while stressing their potential relevance or value.	Encourage elaboration on underexplored arguments by highlighting the significance of addressing those angles.	Pair consensus-building guidance with persuasion about the value of resolving the issue .	Persuade participants of the value of summarizing and linking shared understanding into practical applications.
Participating	Contribute novel perspectives directly without directives.	Add complementary content to underexplored arguments to probe further exploration.	Present own stance to invite co-development of resolutions.	Initiate a partial synthesis with own reflections, leaving space for others to complete.
Delegating	Use pre-defined, minimal scripts at the beginning of each phase to suggest possible participant actions while preserving autonomy.			

Table 2: Summary of findings from the design workshop, including the desired outcomes for each knowledge co-construction phase and the corresponding intervention strategies in different styles. The delegating-style strategy was not derived from the workshop due to its minimal intervention nature.

on design requirements derived from our research motivation and relevant literature, we proposed an intervention logic to guide the agent's behavior toward these objectives. We then designed the agent's architecture, which integrates this logic into a coherent computational framework.

3.3.1 Design Requirements. To derive design requirements (DRs) that are theoretically grounded and empirically relevant, we conducted a narrative literature review [9]. We employed a hybrid search strategy, combining keyword-based retrieval on Google Scholar (e.g., *"knowledge co-construction," "collaborative knowledge building"*) with snowballing from seminal frameworks like IAM [51] and Stahl's model [117]. We targeted empirical studies—ranging from observational analyses of discussion datasets to experimental evaluations of collaborative processes—that investigate the dynamics of knowledge construction. From this pool, we prioritized research situated in asynchronous contexts that identified barriers to effective co-construction. A total of 26 key papers were synthesized to formulate the following DRs:

DR1 Progression Facilitation: Models of knowledge co-construction (e.g., IAM [51], Stahl's model [117]) conceptualize it as a sequence of progressively demanding phases, yet empirical work of asynchronous discussions shows most contributions remain clustered in early stages. Participants typically lack explicit awareness of the discussion's current phase, and sufficient skills to determine when and how to advance the discourse [43, 79]. Therefore, the agent should monitor the evolving discussion and deliver well-timed stimulus, aiming to promote progression towards higher phases rather than forcing all discussions to reach the final co-construction phase.

DR2 Phase Sufficiency: The agent should avoid advancing discussions prematurely, as each phase provides a necessary foundation for the next [117]. Progressing to later stages without fully developing the preceding ones risks undermining collaboration quality: without diverse initiation, later stages lack the breadth of perspectives needed for productive exchange, leading to narrow deliberation [117]; without thorough exploration, negotiation may devolve into shallow or emotional disputes rather than rational debate [104]; without rigorous negotiation, synthesis is prone to superficial consensus, producing fragile rather than durable knowledge [134]. The agent should therefore ensure that each phase meets the sufficiency criteria – defined in the design workshop – before moving forward.

DR3 Agency Preservation: While well-timed support can foster sustained participation [106, 118], overly frequent or prescriptive interventions may erode autonomy, diminishing

intrinsic motivation, ownership of ideas, and creative contributions [33, 110]. The agent should therefore act selectively, offering only the necessary input to sustain productive dialogue while leaving participants sufficient space to steer the discourse and co-construct knowledge on their own terms.

Guided by **DR1** and **DR2**, we propose a high-level intervention logic for advancing knowledge co-construction in asynchronous discussions (Figure 3). At each phase, the agent evaluates whether the discussion meets the sufficiency criteria defined for that stage. If underdeveloped, the agent applies targeted interventions until the desired outcomes are achieved. Once met, the agent initiates strategies for the next phase to foster phase transition, and repeats the process. Specifically, this process accommodates latecomers contributing earlier-stage ideas without suppressing them to enforce a strict forward progression.

3.3.2 Agent Architecture. We implemented an LLM [2] agent whose architecture comprises five core components: *Comment Analyzer*, *Frequency Controller*, *Phase-Sufficiency Evaluator*, *Style Manager*, and *Response Generator* (Figure 3). All numerical thresholds (e.g., intervention intervals) used in this section were empirically derived from pilot trials to ensure smooth discussion flow in our experimental setting. Technical details (e.g., algorithms, prompts, evaluations) are included in the supplementary material.

Comment Analyzer : This component continuously monitors incoming comments and processes them through its submodules to construct a structured understanding of the discussion, which serves as the basis for subsequent interventions.

1. *Phase Classifier* — Consistent with previous empirical approaches [39, 102, 144], we treat each comment as a distinct analytical unit, categorized into one of five categories corresponding to the knowledge co-construction phases (0 = no contribution to knowledge advancement, 1–4 represent Phases 1–4). To build this classifier, we collected data from two strictly moderated communities in Reddit [109], r/AskScience and r/ExplainLikeImFive, as they foster collaborative sense-making and iterative explanation through multi-turn dialogues, aligning well with the dynamics of knowledge co-construction. To ensure comprehensive topic coverage, we utilized the official Reddit API to stratify our sampling across diverse domain tags inherent to these subreddits (e.g., *Economics*, *Social Science*), collecting a balanced number of threads per domain. We prioritized threads with high comment volume to capture the full spectrum of knowledge co-construction phases. Data collection concluded on June 25, 2025, yielding a raw dataset of 13253 comments across 225 threads. To ensure data quality, we removed comments marked as [deleted] or [removed], resulting in a refined corpus of 12587 comments.

Two authors manually coded 1,500 comments randomly sampled from the dataset. They first independently labeled an initial set of 100 comments to align interpretations of phase definitions, then independently labeled the remainder. Disagreements were resolved through discussion; unresolved cases were adjudicated by a third coder. Using the 1500 labeled samples, we fine-tuned a Mistral model [71] with a LoRA adapter [65], achieving an F1-score of 82.6% (train/test split: 70/30). This demonstrated that the classifier generalizes well across varied discussion topics.

2. *Reply Identifier* — This module detects whether a new comment is connected to any earlier comment, either via explicit reply links or implicit references (i.e., addressing earlier content without using the reply function). Leveraging LLM-based semantic matching, it identifies the source comment. All comments and their identified links are stored in a dynamic graph for downstream use.

Frequency Controller: This component regulates intervention timing in line with **DR3**, using two triggers: 1. *Time threshold*: if no user comments for one hour—a threshold established in the pilot to prevent silence without dominating the conversation—the agent intervenes to re-activate participation and prevent prolonged inactivity. 2. *Comment threshold*: if ten new comments have been posted without the current phase reaching sufficiency, the agent intervenes; if sufficiency is reached earlier, the count resets. This mechanism addresses potential phase stagnation while avoiding unnecessary interventions when the discussion is on the right track.

Phase-Sufficiency Evaluator: This module determines whether the current phase has reached sufficiency and, based on that assessment, selects the appropriate phase-specific strategy. In our controlled lab study, the workshop-derived sufficiency criteria for Phases 1–4 were adapted to fit the experimental context (Section 4):

- **Initiation**: The goal of this phase is to elicit a diverse set of initial contributions. Therefore, sufficiency is reached when more than three Phase-1 comments have occurred.
- **Exploration**: The desired outcome is to expand on existing ideas by developing complete argumentative structures. We evaluated this based on the comment graph from *Comment Analyzer*, treating each connected component of Phase-1 and Phase-2 comments as a *'collective argument'* [85]. Drawing from established theories in evaluating collective argumentation [127, 134], each argument's structure is scored using a simplified three-item checklist — presence of a Qualifier, Evidence, and Reasoning — with each item scored 0 or 1. A score of at least 2 marks the argument as complete. Counterarguments within a connected component are evaluated in the same way. The phase is considered sufficient when at least 70% of collective arguments are complete. To assess LLM's reliability in applying these criteria, two researchers independently scored Reddit comment trees, resolved disagreements, and compared the consolidated labels with the LLM's predictions. An F1-score of 86.0% indicated that LLM can reliably evaluate argument completeness in this context.
- **Negotiation**: The sufficiency criterion is that every identified conflict is addressed by corresponding Phase-3 comments and resolved to consensus. Specifically, the agent categorizes consensus into three types beyond simple agreement: (1) *Clarified Disagreement* (where divergence is clear but irreconcilable), (2) *Conditional Agreement* (where different positions hold under specific conditions), and (3) *Full Agreement*. LLM has demonstrated capabilities in consensus detection [58], and we conducted a technical evaluation similar to that used for Phase 2 to verify the performance,

[2]During our experiment (August 2025), we used Gemini 2.5 Flash (hereafter "LLM"), selected for its strong balance of reasoning, latency, and cost-efficiency. As our architecture relies on generic capabilities rather than model-exclusive features, we anticipate comparable performance across other SOTA models.

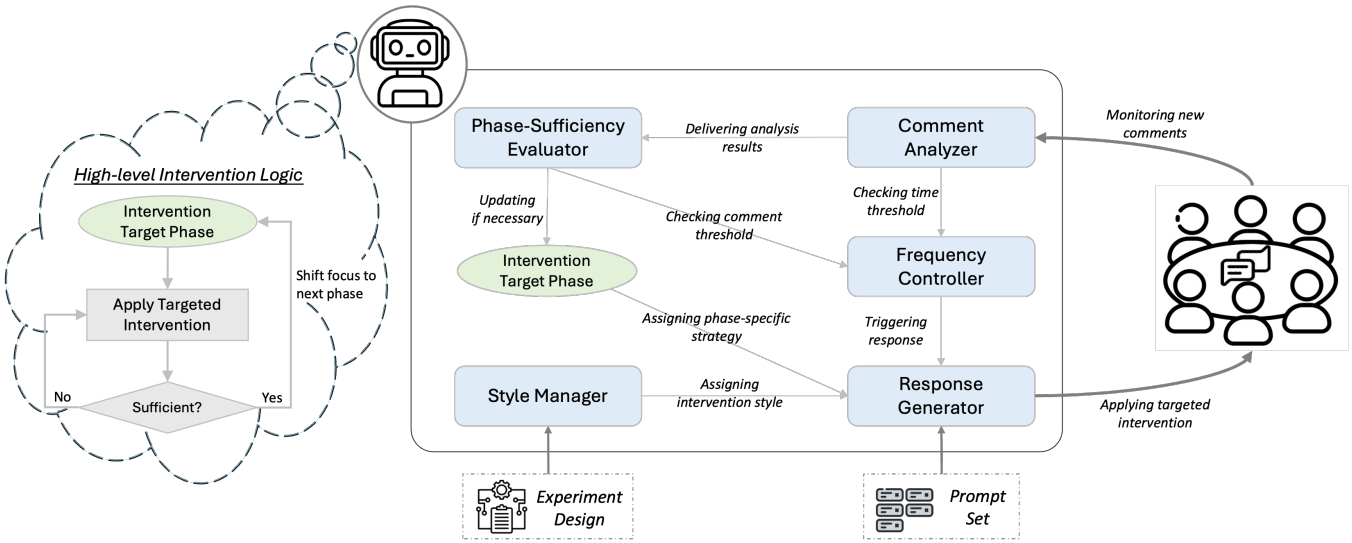


Figure 3: Architecture of the LLM agent implemented in our experiment, detailing its key components and operational logic. The architecture integrates high-level intervention logic to support targeted interventions in multi-party discussions, thereby facilitating the progressive advancement of knowledge co-construction.

where an F1-score of 76.7% demonstrated its effectiveness in detecting whether a conflict has been resolved to consensus.

- **Co-construction**: Here, sufficiency requires that Phase-4 comments collectively summarize all agreements and contain more than three metacognitive statements that either reflect on or apply the collective knowledge.

Style Manager: Each discussion thread in the experiment is assigned a predefined intervention style (see Section 4). This module ensures that generated interventions match the assigned style.

Response Generator: When *Frequency Controller* triggers an intervention, the *Phase-Sufficiency Evaluator* and the *Style Manager* determine which phase-specific strategy to apply and in which style. This module then retrieves the corresponding strategy prompt from a curated set, which were developed based on workshop-derived intervention strategies (Table 2) and refined through iterative prompt-engineering. The selected prompt is passed to the LLM to generate the final response.

Overall, this architecture, integrated with the intervention logic, enables the agent to satisfy the DRs and foster knowledge co-construction in asynchronous discussions.

3.4 Experimental Platform

We developed a web-based experimental discussion platform with the agent integrated, accessible via standard browsers (Figure 4), enabling participants to engage in simulated asynchronous discussions while receiving agent interventions. In this paper, a discussion *thread* denotes the entire unit comprising the initial topic post and all subsequent comments; a *comment* denotes a single user message within a thread.

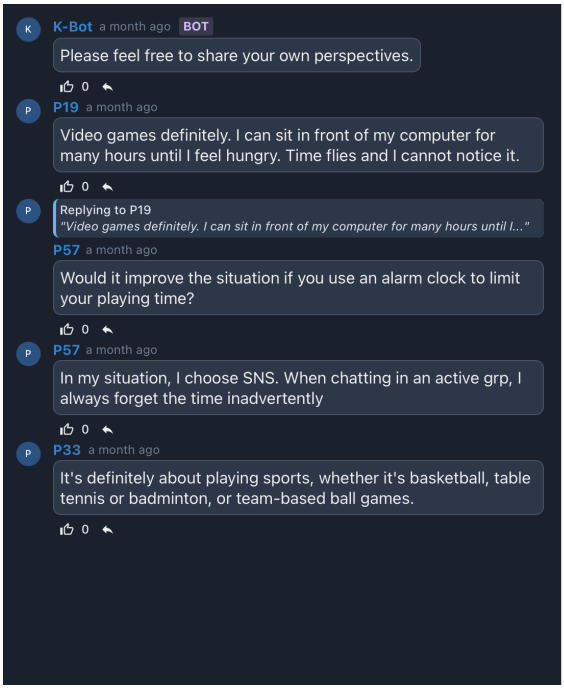
Upon logging in, participants navigated from a dashboard of active threads to expanded views of specific discussions. The discussion interface was designed with reference to Discourse [34], displaying comments within a thread in strict chronological order. Compared with layouts that support nested, tree-structured replies, this structure tends to reduce interactional coherence and result in harder-to-follow discussions [62, 120], which can hinder the progression of knowledge co-construction, providing a suitable context for the agent to exert its influence. Participants could contribute new comments to the discussion via an input box at the bottom of the page or interact with specific comments using "reply" and "like" buttons. To nudge engagement, a notification badge in the top-right navigation bar appeared whenever a participant received a reply. Clicking this notification automatically redirected the user to the relevant comment context. Agent-generated comments were explicitly marked with a `BOT` tag to clearly indicate their automated origin.

4 Experiment Design

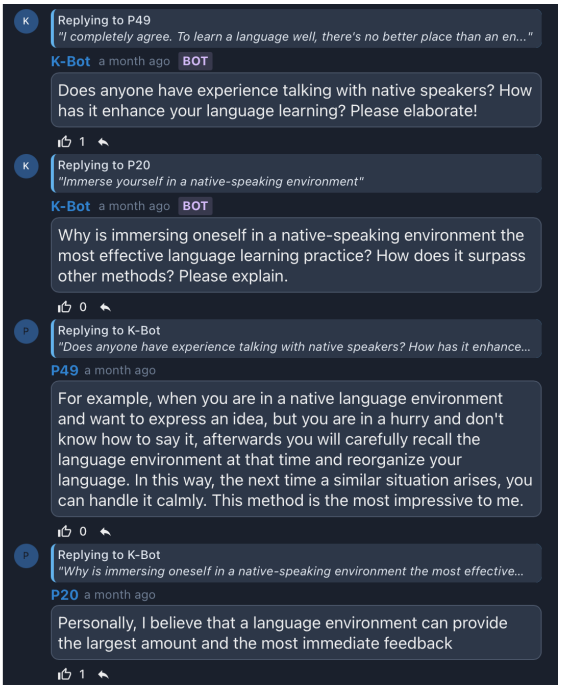
This section details our experimental setup, including conditions, study procedures, and evaluation metrics. We employed a mixed-methods, within-subject design where participants engaged in multiple consecutive discussion threads under different conditions.

4.1 Conditions and Setup

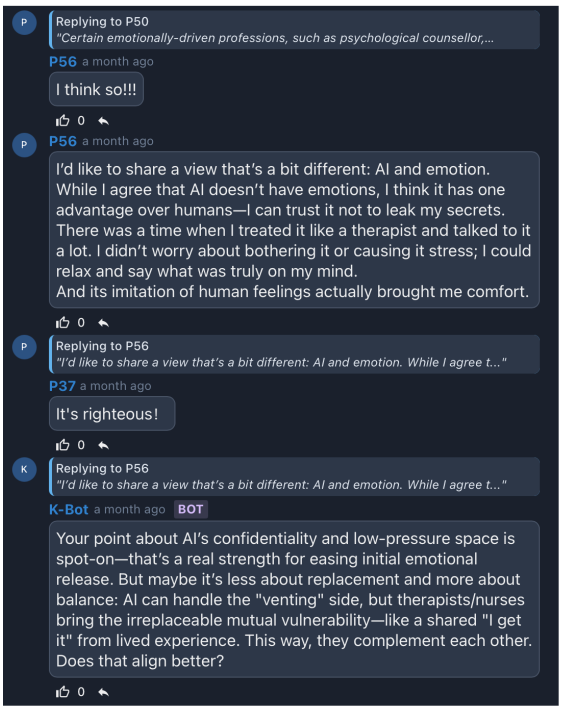
Participants experienced five conditions (four intervention styles and one human-only baseline), with each corresponding to one discussion thread. Studies on online communities [2, 103] indicate that a small fraction of members contribute most of the content, while the majority remain passive ("lurkers") or post infrequently. Thus, we assigned only a small number of active contributors to each thread to simulate the discussion dynamics observed in real-world forums. Pilot study showed that groups of six participants per thread yielded sufficient interaction for meaningful discourse.



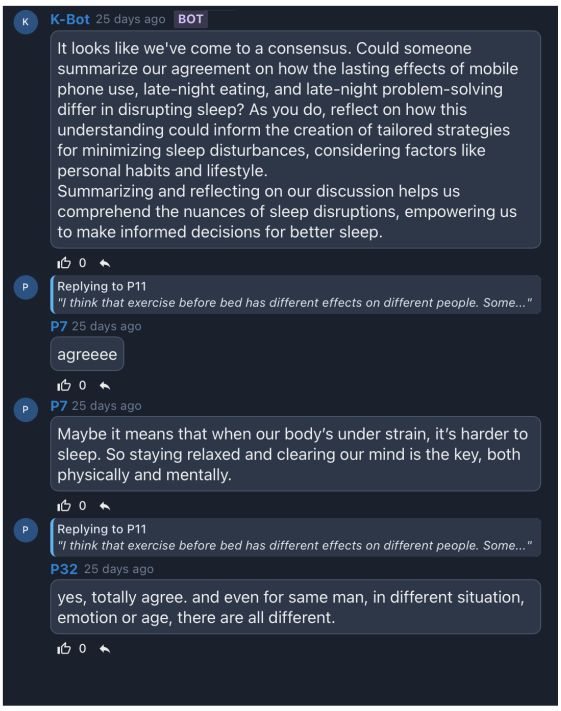
(1) Delegating Intervention for Phase 1



(2) Telling Intervention for Phase 2



(3) Participating Intervention for Phase 3



(4) Selling Intervention for Phase 4

Figure 4: Example screenshots from four threads on the discussion platform used in the lab study, illustrating the interface and instances of agent interventions across different phases and styles. Participants engaged in discussions by posting comments and liking or replying to others. The examples show: (1) a delegating-style intervention tailored to Phase 1, (2) a telling-style intervention tailored to Phase 2, (3) a participating-style intervention tailored to Phase 3, and (4) a selling-style intervention tailored to Phase 4.

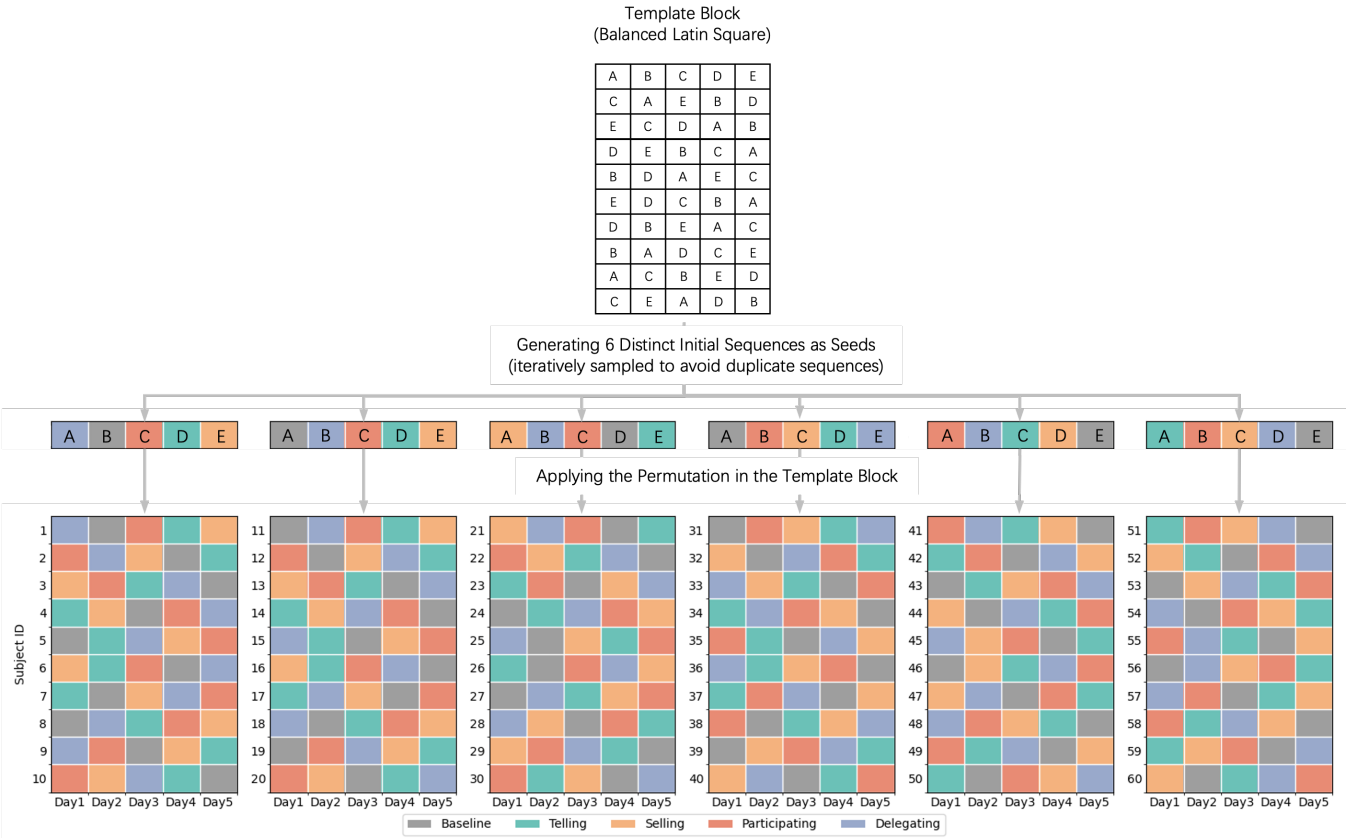


Figure 5: The counterbalancing strategy employed in the study. We iteratively sampled initial sequences and applied the template permutation, discarding duplicates until we obtained six seeds that yielded 60 unique condition sequences.

In total, we collected 50 discussion threads covering different topics. Each thread is discussed by a distinct group of 6 participants. No participant engaged in multiple threads simultaneously to avoid cross-thread confounding effects. Thus, the discussions are conducted sequentially. To simulate asynchronous discussions, participants were given a time window rather than a single-session requirement. Based on pilot trials, we set this duration to one day per thread. Each day, participants could log in at any time to contribute to one thread. At the end of the day, the thread was closed, and participants proceeded to a new topic under the next condition.

We selected five discussion topics, adhering to two criteria: (1) open-endedness, to naturally elicit diverse ideas and encourage negotiation; and (2) accessibility, to ensure that topics require no specialized knowledge, enabling participants to contribute meaningfully as active contributors. An example topic is "What's the most effective way to learn a new language". A full list of topics is provided in the supplementary material.

4.2 Participants and Procedure

We first conducted an a priori power analysis in G*Power [41] based on a repeated-measures ANOVA design with five conditions. Assuming a medium effect size $f = 0.25$, a significance threshold $\alpha = 0.05$, and a statistical power $1 - \beta = 0.95$, the analysis suggested a minimum sample size of 31. With IRB approval, we recruited 60 participants (24 female, 36 male, aged 19-33, with backgrounds in engineering, science, arts and business) through online advertisements, social media, and word-of-mouth. Participants were eligible for the study if they were active contributors of asynchronous discussion platforms. We excluded participants from our previous study to avoid potential bias.

Figure 6 illustrates the experimental procedure. Participants first completed a pre-task survey to collect demographic information and confirm eligibility, followed by a task briefing. They then participated in five sequential one-day discussion threads, each involving a distinct topic. The topic for each day was fixed across the cohort to ensure a shared discussion context, enabling participants to interact within the same thread, with the topics being broadly accessible and of comparable difficulty (Section 4.1). To control for order and carryover effects, we employed a *balanced latin square* design [135]. Given five conditions, a *balanced latin square* block requires 10 sequences to ensure that every condition appears in every position and precedes/follows every other condition equally often. To maximize interaction diversity and ensure participants encountered different peers everyday, we generated six mutually disjoint sets of these blocks. Specifically, we selected six distinct initial sequences as seeds and applied the *balanced latin square*

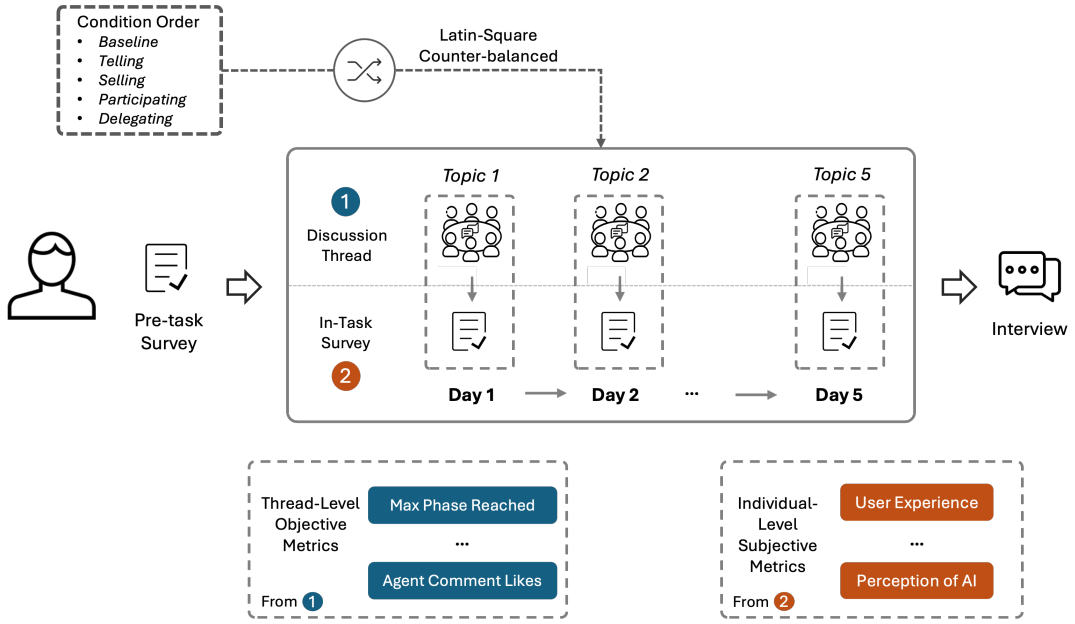


Figure 6: A within-subject study with 5 conditions (baseline + 4 intervention styles). We used a counterbalanced design to avoid order effect. Participants completed discussion threads for the assigned condition, as well as a pre-task survey, in-task surveys, and a follow-up interview.

permutation to each, such that there were no duplicates among the final 60 condition sequences—one for each participant (Figure 5). This design ensures that: (1) every condition appeared equally often on each topic, effectively neutralizing topic-specific biases; and (2) no two participants followed the exact same trajectory, preventing fixed subgroups.

Following each daily discussion, participants completed an in-task survey to provide condition-specific feedback. Upon completing all threads, participants took part in a semi-structured interview to explore their knowledge co-construction experiences, perceptions of different styles, and underlying rationales behind their behaviors. Participants were compensated at the local standard hourly rate.

4.3 Measurements

We evaluated each RQ using thread-level objective metrics and individual-level subjective metrics. All in-task surveys used a 7-point Likert scale; specific items are provided in the supplementary material.

4.3.1 RQ1: How do AI intervention styles influence knowledge co-construction progression? At the thread level, *Max Phase Reached* measured the highest phase achieved, serving as an indicator of the agent's capacity to facilitate progression rather than a requirement that every thread must attain the final phase. *Sufficiency Criteria* indicated whether the thread met predefined requirements for each phase (Table 2). Two researchers independently coded these metrics based on the definitions in Section 3, and resolved discrepancies through discussion. At the individual level, participants rated the discussion's *Smoothness* [7], *Effectiveness* [75, 76], *Depth* [89], and *Sufficiency* [104].

4.3.2 RQ2: How do intervention styles influence participant perceptions of the AI agent and their user experience? Thread-level indicators included the total *number of likes received by the agent's comments* and the *number of human replies to the agent*. Subjective measures captured participants' perceptions of the agent's *Helpfulness* [20, 95], *Appropriateness* [86, 92], *Intrusiveness* [35], and *Social Presence* [12]. User experience was assessed through NASA Task Load Index [55, 143], specifically *Mental Demand [90], Effort [93],* and *Frustration* [95] following existing research, as well as *Perceived Complexity* [91, 94] and *Agency* [67].

4.3.3 RQ3: How do intervention styles affect human-human interaction? The primary thread-level metric was the *number of human replies to other human participants*. Perceptions of interpersonal connection were captured through the Inclusion of Other in the Self (*IOS*) scale [4, 66] and perceived *Mutual Awareness* [107].

4.4 Analysis Methods

We employed a mixed-methods approach for data analysis. Quantitatively, for individual-level measures, Shapiro–Wilk tests [112] indicated violations of normality. Consequently, we applied the Friedman test [46], a non-parametric alternative suitable for our within-subjects design. Significant main effects were followed by Bonferroni-corrected Wilcoxon signed-rank post-hoc tests. We report test statistics, adjusted p-values, and effect sizes. For thread-level metrics, we employed Generalized Linear Mixed Models (GLMM) [19] to model non-normal distributions without transformation and to account for data dependencies. Thread metrics were specified

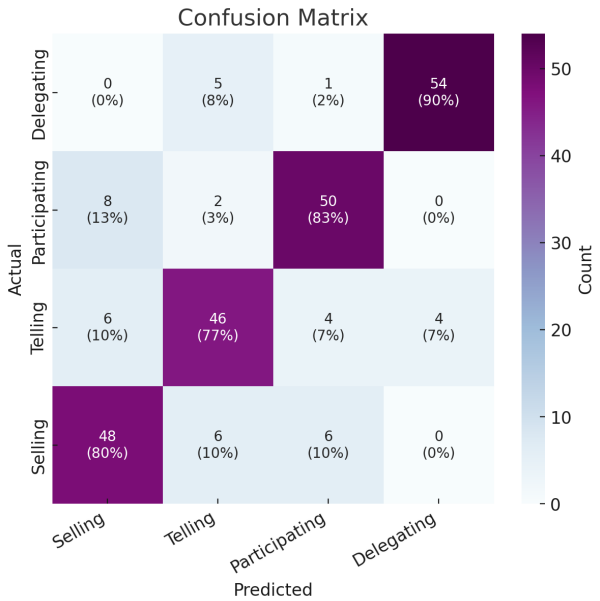


Figure 7: Confusion matrix comparing participants' perceived intervention style of each agent with the actual style implemented. Darker cells indicate higher classification counts.

as dependent variables, Condition as a fixed effect, and Day as a random intercept to control for day-specific variability. We specified model distributions based on variable type: binomial for binary outcomes (e.g., Phase Sufficiency), negative binomial for count variables (e.g., Agent Comment Likes), and cumulative link models for ordinal variables (e.g., Max Phase Reached). Upon observing significant main effects, we conducted pairwise comparisons using estimated marginal means with Benjamini–Hochberg (BH) correction [11]. We report model estimates, standard errors (SE), and p-values, alongside effect sizes presented as Odds Ratios (OR) for binary/ordinal outcomes and Rate Ratios (RR) for count data. Across analyses, items for RQ1 and RQ3 were compared across all five conditions, while RQ2 focused on four intervention groups.

Qualitatively, we conducted a thematic analysis [17] of the semi-structured interview transcripts, grounded in each RQ [53, 54]. First, audio recordings were automatically transcribed via Zoom. Two researchers reviewed all transcripts to ensure fidelity and got familiar with the data. Second, they performed a structural categorization by mapping the interview segments to the corresponding RQs. Third, within these RQ-aligned categories, the two coders independently coded all the transcripts using ATLAS.ti [5], and frequently met during the analysis to reconcile discrepancies through negotiated agreement until consensus was reached. After key themes emerged from the data, they were refined through iterative team discussions. The final themes and codes are summarized in Figure 11, which systematically maps specific insights to corresponding RQs, with detailed interpretations and representative quotes provided in Section 5.

5 Results

This section presents the results for each RQ, integrating statistical analyses with qualitative insights (summarized in Figure 11) to offer a comprehensive view. For brevity, we capitalize style names to refer to either agents exhibiting that style or the corresponding condition; for example, "Selling" denotes selling-style agents or the selling condition.

5.1 Manipulation Check

To validate our manipulation, we compared participants' judgments of each agent's intervention style with the style that the agent was programmed to enact, as shown in Figure 7. Overall, 82.5% of classifications matched the intended style, suggesting that participants generally recognize the correct style. A Chi-square test confirmed that the observed classification frequencies significantly differed from chance ($\chi^2(9) = 432.58, p < .001$), indicating that the agents were reliably perceived in line with their designated styles. This verified that our manipulation of the independent variable was effective, ensuring that our participants perceived, understood, and reacted to the four intended agent styles as expected.

5.2 RQ1: How do AI intervention styles influence the progression of knowledge co-construction?

We first compared discussions with AI intervention (four styles) against the no-agent baseline. We then examined differences among the styles in shaping the overall knowledge co-construction process, followed by a phase-level analysis of how each style facilitated progression.

Metrics	Baseline	Telling	Selling	Participating	Delegating
Max Phase Reached (Mean ± SD)	2.10 (0.32)	2.80 (0.63)	2.90 (0.88)	2.60 (0.52)	2.20 (0.42)
Phase 1 Sufficiency (%)	90%	100%	100%	100%	100%
Phase 2 Sufficiency (%)	10%	80%	60%	70%	20%
Phase 3 Sufficiency (%)	0	30%	30%	30%	0
Phase 4 Sufficiency (%)	0	0	0	0	0
Agent Comment Likes (Mean ± SD)	NA	0.90 (0.88)	0.60 (0.70)	2.20 (1.48)	0.20 (0.63)
# Human Replies to Agent (Mean ± SD)	NA	4.30 (2.41)	4.00 (2.36)	4.00 (1.83)	0.50 (0.85)
# Human–Human Replies (Mean ± SD)	6.40 (2.32)	2.90 (3.81)	3.90 (2.69)	4.10 (2.60)	6.10 (3.41)

Table 3: Descriptive statistics of thread-level measures across the five experimental conditions. The metrics *Agent Comment Likes* and *Number of Human Replies to Agent* are reported only for the four intervention-style conditions, as the baseline condition does not include an agent.

5.2.1 Discussion Dynamics with Agent Intervention. Table 3 presents descriptive statistics of the thread-level measures, and Table 4 reports pairwise comparisons (only dyads with at least marginal significance are shown for brevity; full results are provided in the supplementary material). As shown in these tables, all intervention conditions other than Delegating was associated with markedly deeper progression of knowledge co-construction than the baseline. The average *Maximum Phase Reached* increased by 23.8–38.1% (Table 3); pairwise comparisons reached marginal significance with substantial odds ratios (ORs: 0.04 – 0.09) (Table 4), suggesting practically meaningful improvements despite not meeting conventional thresholds. Baseline discussion typically plateaued in the early stages, with only rare comments reaching Phase 3 or beyond, whereas agent-intervened threads consistently advanced further through the phases. Phase-sufficiency analysis reinforces this divergence. While 90% of baseline groups met sufficiency in Phase 1, only 10% reached sufficiency in Phase 2. In comparison, Telling, Selling, and Participating conditions yielded *Phase 2 sufficiency* in 60–80% of cases, reaching marginal significance against baseline yet showing substantial differences in odds (ORs: 0.03–0.07). (Table 4). Moreover, these three styles uniquely achieved Phase 3 sufficiency in about 30% of threads, an outcome never observed without an agent. This indicates that agents not only increased the likelihood of advancing deeper but also strengthened earlier phases as foundations for further progress. These findings align with the subjective evaluations (Figure 8). Compared with the baseline, participants perceived significantly higher levels of *Depth* and *Sufficiency* of knowledge co-construction under Telling, Selling, and Participating interventions, while Delegating showed no significant difference.

Interview feedback revealed why the agent was able to advance knowledge co-construction. First, as a **low-friction interlocutor**, the agent lowered the social cost of speaking up, making it easier to post the *"first sentence"* (P37, P42, P59). As one participant noted, *"When discussing with AI, I didn't feel any burden or that I was inconveniencing others"* (P37). This consequently expanded the pool of visible contributions at the Initiation stage, laying the groundwork for subsequent enrichment and deepening of knowledge. Once participants began contributing, the agent provided **validation and psychological safety**. Even brief replies by the agent conveyed that those contributions were seen and valued, alleviating the perceived risk of being ignored and encouraging participants to elaborate. *"When the bot replied, I felt more at ease… even a bot's response made me feel my comment was acknowledged"* (P17). Another added, *"Receiving [bot's] reply… prompted me to explore the topic more deeply"* (P40). Such reassurance not only offered emotional comfort but also made it seemingly worthy of expanding tentative ideas into more detailed and substantive contributions. When discussions slowed down, the agent acted as a **momentum catalyst** that reactivated stalled threads, provided footholds for others to pick up and continue, and kept ideas flowing. As P42 observed, *"When progress was slow, AI motivated us and kept the discussion going"*. Others emphasized that without such prompts, *"many continuations wouldn't have happened"* (P12). Beyond maintaining rhythm, the agent also functioned as a **perspective expander**. It surfaced overlooked angles and concrete details (P12). Engagement was strongest when prompts resonated with existing thinking—*"I respond when the bot's reply connects with my cognition"* (P17). The agent also introduced alternative viewpoints, nudging participants away from one-sided statements and toward considering multiple perspectives. *"Humans don't usually present both pros and cons; the bot took that role and guided me to consider the opposite view"* (P4). Meanwhile, not all participants embraced the agent as a true member of the group, as P56 put it, *"I wouldn't reply to a bot; I'd rather talk to people"*. This reluctance varied among different agent styles, a contrast we present further in Section 5.2.2.

5.2.2 Differential Impacts of Intervention Styles. Having examined the effect of agent presence in general, this section compares how the four styles differentially shaped knowledge co-construction. Overall, both task-oriented styles—Telling and Selling—marginally outperformed Delegating in *Max Phase reached*, and Telling retained this advantage for *Phase 2 Sufficiency* (Table 4). Despite these statistical similarities, participants perceived the two styles differently. Selling was appreciated by some for its elaborative explanations, which they felt **stimulated idea generation and deeper reasoning**. As P23 explained, *"These explanations broadened my thinking*

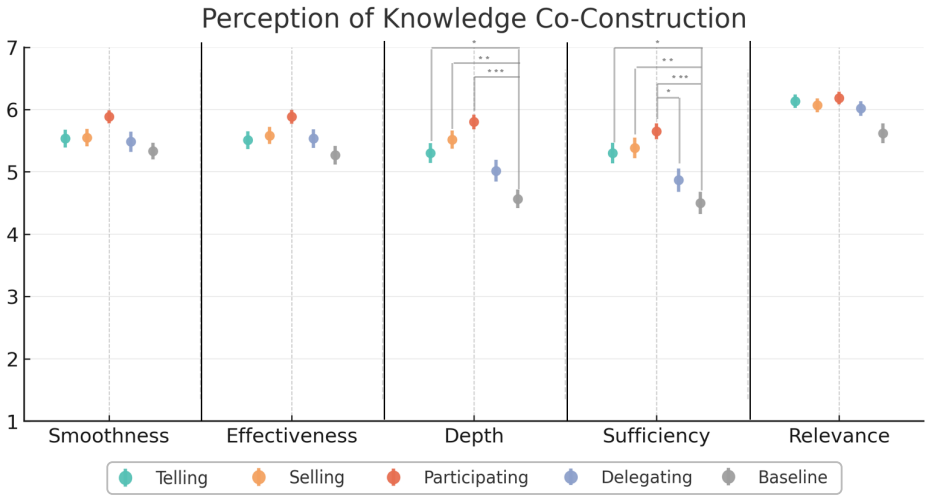
Comparison	Estimate	SE	p-value	Sig.	OR/RR
Max Phase Reached					
Baseline – Telling	2.92	1.22	0.083	+	0.05
Baseline – Selling	3.14	1.26	0.083	+	0.04
Baseline – Participating	2.40	1.21	0.093	+	0.09
Delegating – Telling	2.13	1.00	0.084	+	0.12
Delegating – Selling	2.34	1.05	0.084	+	0.10
Phase 2 Sufficiency					
Baseline – Telling	-3.58	0.04	0.052	+	0.03
Baseline – Selling	-2.60	0.09	0.071	+	0.07
Baseline – Participating	-3.04	0.06	0.052	+	0.05
Delegating – Telling	-2.77	0.07	0.052	+	0.06
Delegating – Participating	-2.23	0.11	0.071	+	0.11
Agent Comment Likes					
Delegating – Telling	-1.50	0.17	0.081	+	0.22
Delegating – Participating	-2.40	0.07	0.007	**	0.09
Participating – Telling	0.89	0.97	0.048	*	2.44
Participating – Selling	1.30	1.70	0.014	*	3.67
# Human Replies to Agent					
Delegating – Telling	-2.15	0.06	<.001	***	0.12
Delegating – Selling	-2.08	0.06	<.001	***	0.13
Delegating – Participating	-2.08	0.06	<.001	***	0.13
# Human–Human Replies					
Baseline – Telling	0.79	0.49	0.004	**	2.20
Delegating – Telling	0.74	0.47	0.005	**	2.10

Table 4: Pairwise comparison results for thread-level measures. *Only dyads with at least marginal significance (p <.10) are included.* Reported values include the model estimate, standard error (SE), p-values (+: .050 < p <.100, *: p <.050, **: p <.010, ***: p <.001) adjusted using the Benjamini–Hochberg (BH) correction procedure,, and effect sizes (odds ratios for ordinal/binary variables and rate ratios for count variables).

and made it easier to follow the agent's reasoning." However, others found it **overly verbose, obscuring key guidance** (P31, P52). Telling, in contrast, was praised for its clarity and conciseness. As P60 noted, *"**Providing a clear guide to follow** makes the process much easier"*. Yet, this style also had drawbacks. A few participants questioned its impact, noting that its brevity was easy to ignore, and its **directness discouraged participation** (P17-19, P56). As P17 explained, *"It urged me like a task to finish, which suppressed my willingness to contribute"*." Notably, a group of participants (13/60) explicitly **resisted engaging** with either style, perceiving them as *"non-human authorities"*. *"I don't want to spend energy on an AI"*, said P59.

Compared with Delegating, Participating approached a significant improvement in *Phase 2 Sufficiency* (Table 4) and was perceived as fostering greater *Sufficiency* in discussion (Figure 8). Despite its low task-orientation, it achieved outcomes comparable to more directive styles. Many participants (37/60) attributed this to its resemblance to a typical human member, often treating the agent as a peer. Some even reported forgetting that it was a bot, *"It sounded so much like a person that I just replied instinctively"* (P26). This peer-like quality encouraged **leadership by example** (P10, P23) and created **an equal, immersive interaction that fostered reciprocity**, generating a virtuous cycle of contributions (P8, P26, P53). However, some cautioned that its guidance was too subtle, risking stagnation when stronger direction was needed (P39, P42). Others noted that when the agent presented well-formed viewpoints, it paradoxically **reduced their inclination to contribute further** (P21, P59). As P21 noted, *"It felt so smart that I didn't need to add anything"*.

Delegating showed no significant differences from the baseline across all metrics (Table 4), indicating minimal impact on discussion progression. Most participants (44/60) reported largely ignoring its



Comparison with *	Adj. p	Effect Size	Z
Depth vs.	.014	.41	3.20
Depth vs.	.001	.51	3.93
Depth vs.	<.001	.65	5.02
Sufficiency vs.	.020	.40	3.09
Sufficiency vs.	.004	.46	3.55
Sufficiency vs.	<.001	.58	4.50
Sufficiency vs.	.032	.38	2.94

Figure 8: Participants' perceptions of knowledge co-construction across conditions. Error bars indicate 95% confidence intervals (*$p < .050$, **$p < .010$, ***$p < .001$). For significant pairwise comparisons, test statistics and effect sizes are reported. For all items, a higher score indicates better performance.

interventions due to minimal presence, as P37 remarked, *"It didn't affect our discussion"*.

5.2.3 Phase-Specific Effects of Intervention Styles. Beyond overall effects, we examined how each style influenced specific phases of knowledge co-construction through qualitative insights.

Initiation. In this phase, participants introduce perspectives that set the stage for subsequent discussion. Telling and Selling often functioned as **idea seeders**, lowering the entry cost of participation and diversifying discussion starting points. As one participant explained, *"It gave me a seed on something unfamiliar but interesting. I looked it up, reflected, and shared on the forum; without the bot's comment, I wouldn't have done so"* (P17). However, Selling's additional explanations sometimes imposed a **premature interpretive lens**, which participants felt could constrain the openness of thought and channel ideas into narrower tracks (P36). By contrast, Telling was praised for offering concise prompts without persuasive framing, thereby *"providing direction while leaving room for user elaboration"* (P36, P53). Participating contributed its own viewpoints to **broaden discussion** (P3, P21); when its perspective deviated from participants' existing cognition, it often **triggered deeper reflection**. *"When the bot's view was inconsistent with mine, it pushed me to think more deeply about why it was not the same. Unlike with humans, I felt no pressure to oppose it, so I could easily strengthen my own position from the opposite direction"* (P4). Delegating mainly **broke initial silence** through a simple reminder (P40, P42), without substantive direction to shape early-stage contributions.

Exploration. In this phase, participants expanded previous contributions into more complete arguments. Telling and Selling were often described as offering a ***"template"* for enriching arguments**, triggering memories of relevant experiences or knowledge participants might not have otherwise recalled (P43). For Selling, the added persuasion prompted more supplementary comments since it **mitigated the worry of *"contributing something meaningless"***(P20) through explicit justification of why a point merited further elaboration. Yet the same detailed justifications could also make participants feel *"compelled to respond"* as if the bot's reasoning left little room to disengage or disagree (P3, P25, P59). Some participants reacted with **avoidance behavior** — skimming or ignoring the intervention, withholding replies, or posting perfunctory, low-effort responses merely to close the loop (P25, P59). As P59 noted, *"It explained so thoroughly that I felt obliged to reply, almost as if it was forcing me to accept its instructions"*. This avoidance might help explain Selling's lower rate of *Phase 2 sufficiency* compared with Telling and Delegating (Table 3). One typical case was S04 (Figure 9): despite five consecutive Phase-2 interventions by Selling, subsequent replies were either absent or superficial, with only one substantive Phase-2 comment. By contrast, Delegating was sometimes appreciated for **providing minimal but timely nudges** that prompted transition to Phase 2 (P16, P51). Several clear examples emerged, such as D08–10 in Figure 9, where comments were mostly in Phase 1 before the Delegating prompt, but afterward shifted toward Phase 2. From the interview, Participating showed no distinct effect in this phase, with contributions resembling its overall impact.

Negotiation. In this phase, participants worked toward resolving conflicts. All three styles—Telling, Selling, and Participating—were valued for **linking disparate viewpoints** and making participants notice perspectives they might otherwise have overlooked (P2, P44). Telling and Selling provided **actionable entry points for negotiation**. *"They gave me angles to negotiate on, making it clearer how to start and synthesize previous comments"* (P33, P42). Participating operated differently, **carrying a clear stance on reconciling differing viewpoints**. This stance served as a concrete reference for further negotiation (P42, P58) and provoked deeper involvement—*"its words were quite absolute, which made me want to negotiate with it"* (P56)—while others felt compelled to expand

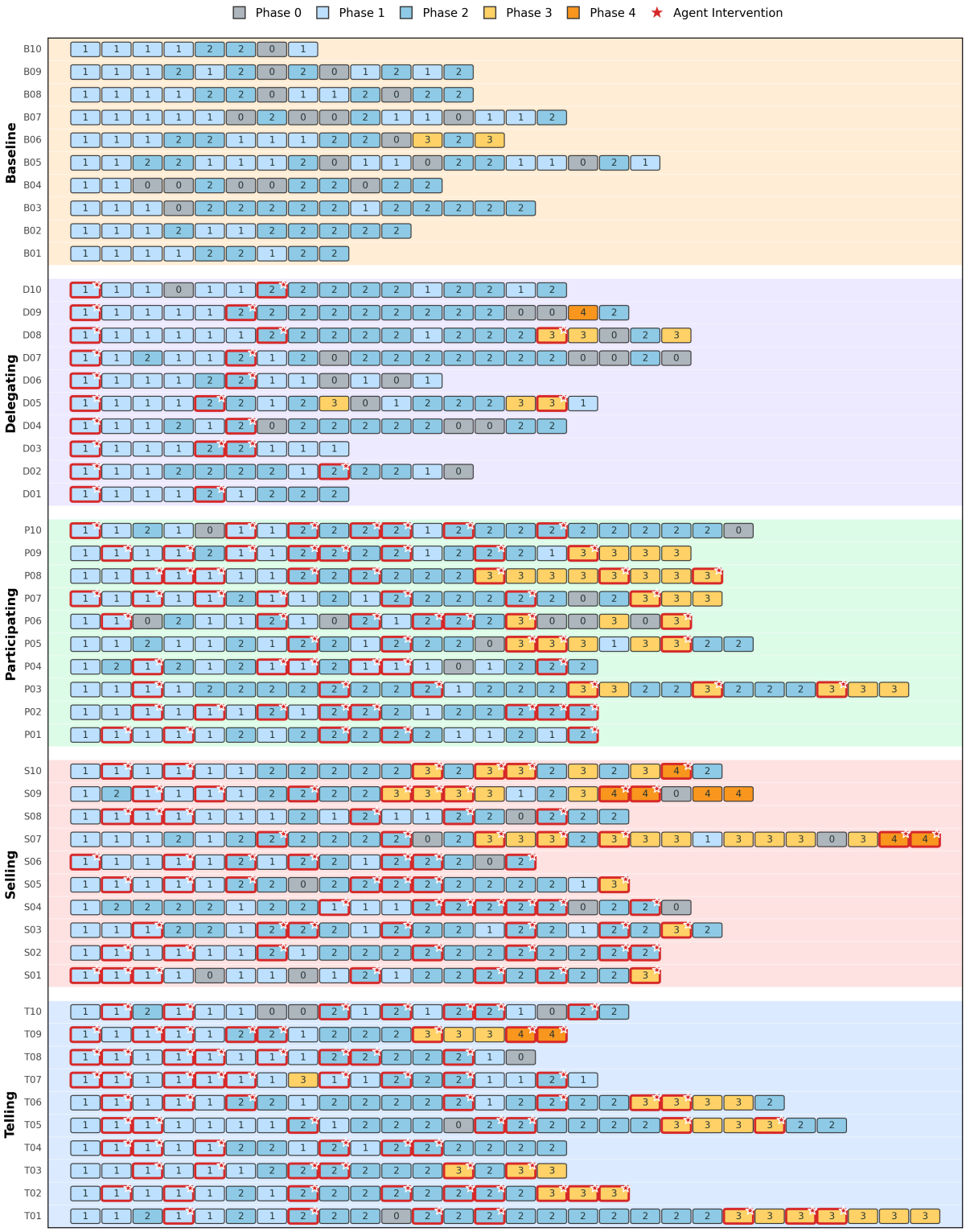


Figure 9: Chronological comment sequences for each thread across different conditions, showing the process of knowledge co-construction. Each cell represents a comment, with the number indicates the phase it is categorized into through manual coding. Red-bordered cells with a red star denote agent interventions, where the number indicates the targeted phase. Colors are also applied to encode phases.

the discussion because *"it wasn't necessarily the most authoritative answer"* (P17). By simulating a realistic discussant, it prompted participants to **articulate counterarguments and refine their positions**. *"when it challenged my stance, it naturally pushed me to respond"* (P18). This dynamic made Participating particularly effective for preventing stagnation in Phase 2 and facilitating transitions into Phase 3. As seen in Figure 9, nearly all threads with a Phase-3 intervention by Participating received follow-up Phase-3 comments. By contrast, in threads like S09 or T02, Telling or Selling often required multiple consecutive Phase-3 interventions to elicit even a single negotiation comment, and sometimes elicited none.

Co-Construction. In this phase, participants synthesize and reflect on the group's shared knowledge. In our experiment, Phase-4 interventions appeared only in four threads under the Telling and Selling conditions—consistent with their task-oriented nature—and elicited just one Phase-4 response. Many users viewed summarizing and reflecting as responsibilities for senior members or AI itself, rather than ordinary members (P2, P23). Summarization was seen as an evaluative act that implicitly claims authority over the group's knowledge, creating **psychological barriers** as participants worried about overstepping or being judged on the quality of their synthesis (P42). Participants also noted that producing a synthesis carried substantially **higher cognitive and emotional demands** than the lighter, incremental contributions of earlier phases, further discouraging engagement (P51). Consequently, directive behaviors that explicitly asked participants to summarize or reflect, whether through Telling's blunt prompts or Selling's persuasive framing, were largely ineffective in transitioning discussions into the co-construction stage.

5.3 RQ2: How do intervention styles influence human perceptions of the AI agent and their user experience?

5.3.1 Perception of Agent Styles. Participants' perceptions of the agents varied across styles, with multiple pairs showing notable differences in both subjective and objective measures. Delegating elicited significantly fewer human replies compared with the other three conditions (Table 4). Subjectively, it was consistently rated as significantly less *Helpful*, less *Appropriate* (except vs. Selling), and lower in *Social Presence* (except vs. Telling) than all other styles. This aligns with its minimal, mechanical, and low-frequency design. Participants often described it as *"system-generated tips"* rather than a genuine member (P13, P34, P41). They acknowledged such **Minimalist Nudge** as **less helpful or socially engaging** but also **less disruptive**; several even noted that this minimalist presence **reduced the *"feeling of being spied"*** compared with other agents (P8, P21).

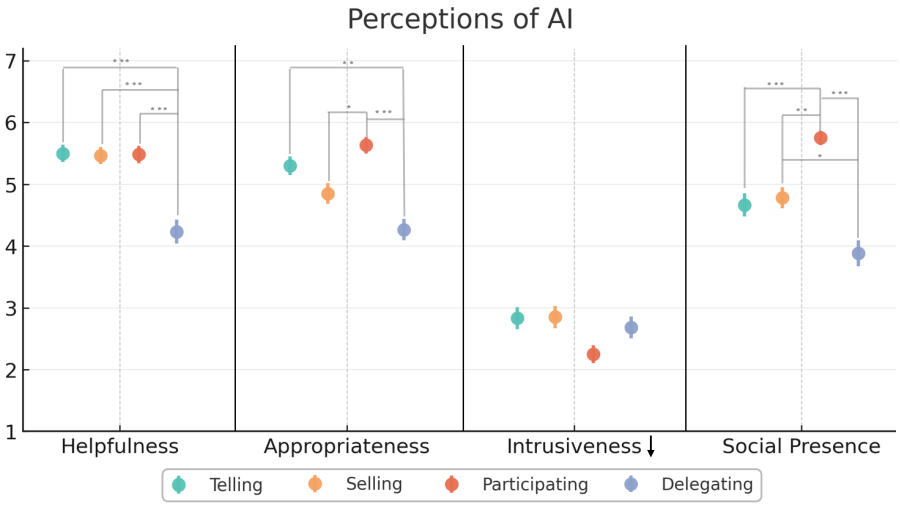
Participating, by contrast, received significantly more *Likes* than the other three styles (Table 4). It was rated more *Appropriate* than Selling and higher in *Social Presence* than both Telling and Selling (Figure 10). As discussed in Section 5.2.2, this can be attributed to its perceived humanness. Participants tended to treat it as a **Peer Collaborator** offering equal and natural interaction. P23 noted that, *"It felt like it was discussing with us rather than leading us [unlike Selling or Telling]."*. Although *Human Reply Counts* did not differ significantly between Participating and Telling/Selling, interviews revealed a **greater willingness to engage** with Participating's comments because they resembled human dialogue (P33, P43), whereas responses to Selling/Telling often felt more like compliance with instructions (P8, P46). This perceived humanness also **softened aspects that were viewed negatively** in other styles. For instance, several participants complained that Selling or Telling sometimes replied selectively, which felt unresponsive —*"as if the AI should address everything"* (P12). By contrast, when the Participating agent did not respond to a comment, participants interpreted it as a natural human-like choice rather than a system flaw. P49 explained that, *"Even if it skipped my comment, I saw it as a natural reaction, not an AI defect."* Similarly, while some participants (14/60) noted that Selling's long message created cognitive load, they tolerated Participating's - *"Even if its messages were long, they didn't feel tiring to read"* (P31).

Selling and Telling did not differ significantly in quantitative measures, but participants' perceptions diverged. Telling was often described as resembling an **Authoritative Boss**, distant, and pushy, which sometimes provoked negative emotional reactions (P17, P37). Some explicitly called its tone *"cold"* or *"rigid,"* highlighting its **mechanical feel** (P17, P24). Selling, by contrast, was frequently likened to a **Guiding Teacher**. Participants felt discussions with its interventions resembled classroom activities, sometimes supportive but also at times verbose and overly didactic (P18, P26, P52). While some appreciated its ***"kind and approachable"* tone** (P15), others found it ***"long-winded"* and heavy-handed** (P25).

5.3.2 User Experience. As shown in Figure 12, there were no significant differences across the five conditions in perceived *Effort, Frustration, Complexity*, or *Agency*. However, participants reported significantly higher *Mental Demand* for Selling compared with Baseline. As elaborated in Section 5.2.3, Selling's detailed, persuasive explanations made participants feel they could not skip them, creating **a sense of obligation**—*"hard to reject"*—that increased emotional burden. The **relatively long messages** further contributed to this demand. Regarding Telling, participants noted that its **conciseness created less psychological burden**; its comments could be *"taken or left"* and thus were not perceived as disruptive (P8). However, when Telling addressed sensitive topics, its **bluntness produced discomfort**. P39 recalled, *"especially when it brought up real-life jobs, it felt too specific and uncomfortably close."* Participating was valued for its non-directive style, which neither enforced agreement nor imposed constraints, fostering **a sense of freedom**: *"I felt free to engage with it or not, without pressure"* (P56).

5.4 RQ3: How do intervention styles affect human-human interaction?

While subjective ratings of human–human interaction showed no significance (Figure 12), thread-level data revealed that Telling elicited significantly fewer human–human replies compared with both Baseline and Delegating (Table 4). This result resonates with a broader concern raised in interviews. Some participants (17/60) mentioned that the agent's interventions—particularly when manifesting as concise and clear comments like Telling's—**shifted attention toward human–agent rather than human–human exchanges**. As P60 noted, *"People chose to interact with the bot,*



Comparison with *	Adj. p	Effect Size	Z
Helpfulness vs.	<.001	.55	4.28
Helpfulness vs.	<.001	.53	4.14
Helpfulness vs.	<.001	.56	4.31
Appropriateness vs.	.005	.43	3.32
Appropriateness vs.	.043	.35	2.69
Appropriateness vs.	<.001	.61	4.70
Social Presense vs.	<.001	.51	3.96
Social Presense vs.	<.001	.77	5.98
Social Presense vs.	.008	.42	3.22
Social Presense vs.	.035	.36	2.76

Figure 10: Participants' perceptions of agents across styles. Error bars represent 95% confidence intervals (*$p < .050$, **$p < .010$, ***$p < .001$). For significant pairwise comparisons, test statistics and effect sizes are reported. Unless marked with ↓ (indicating lower is better), higher scores denote better performance.

distracting from interactions with humans". When the AI replied to a human comment, subsequent contributions often addressed the AI instead of the original commenter (P21, P40). In such cases, the AI tended to substitute for, rather than mediate, peer interaction, leaving participants with fewer opportunities to engage directly with one another (P54). Yet, this disruptive effect was not absolute. In many other instances, users viewed agents as stimulants for interaction. Several participants felt that agents activated discussion and created more opportunities for mutual replies (P18–20, P59). Moreover, when agents (except Delegating) replied to human comments, participants perceived higher quantity and quality of peer exchanges, for two main reasons: (1) such replies **drew attention to comments** that might otherwise have been overlooked, prompting others to engage (P30, P31, P41); and (2) by elaborating on prior comments, the agent **helped others better understand them**, leading to more substantive responses. As P46 explained, *"At first some comments were short and unclear, but after the AI's reply I understood them and wanted to respond seriously."* Some also noted that the **recognition signaled by bot replies** also encouraged more idea exchange (P7, P60). As P60 put it, *"When it replied to me, I felt acknowledged, and became more willing to interact with others."*

Participants also highlighted style-specific insights. For Phase-2 interventions of Telling and Selling, directives or persuasion targeted at specific human comments **lowered the barrier for some to reply** to those messages (P3–5, P17, P33). As P17 explained, *"I usually feel awkward supplementing others' comments because I fear criticism, but with the bot's directive it felt natural"*. In such cases, participants attributed their decisions to interact with others to the bot, which implicitly reduced their perceived risk of being criticized (P5, P33). Telling sometimes reinforced human–human exchanges by being perceived as more detached from the human discussants, subtly **strengthening an *"ingroup–outgroup"* dynamic** [121]. As P24 put it, *"The AI was on the opposite side, so we paid more attention to our human-to-human communication."* By contrast, Selling *"resembled a teacher initiating classroom activities"* (P22); participants **treated its interventions as assignments to complete**, rather than opportunities for peer dialogue, which consequently diminished human–human exchange. Participating was perceived as fostering **more *"natural"* and *"smoother"* communication atmosphere** (P53). Moreover, the agent's own viewpoints occasionally **triggered collective response**—opposing or debating its views—which some felt strengthened group cohesion: *"When we all refuted it together, our connection felt closer"* (P37). Delegating was appreciated as a *"hands-off"* style that left interactions undisturbed. However, participants noted interactions often remained shallow in this condition, consisting mainly of repetition or simple agreement (P8, P39).

6 Discussion

Our evaluation demonstrates that LLM agents equipped with phase-sensitive strategies can reliably advance asynchronous discussions toward deeper phases of knowledge co-construction. Task- and relationship-oriented styles each exhibit distinct strengths and limitations. Drawing on these findings, we first highlight the effectiveness of the proposed process-orchestrated design, then discuss its limitations and generalizability. We next examine the tensions between different styles and explore potential resolutions, and finally propose key design considerations for future research.

6.1 Process-Orchestrated, Phase-Sensitive Agent Design

The agents adopted a simple but effective intervention logic: **ensure sufficiency in the current phase before explicitly pivoting to the next phase**. This differs from prior work that optimizes for single-phase behaviors. For example,the agent proposed by Ito et.al. [69] prompts for elaboration and clarification within the

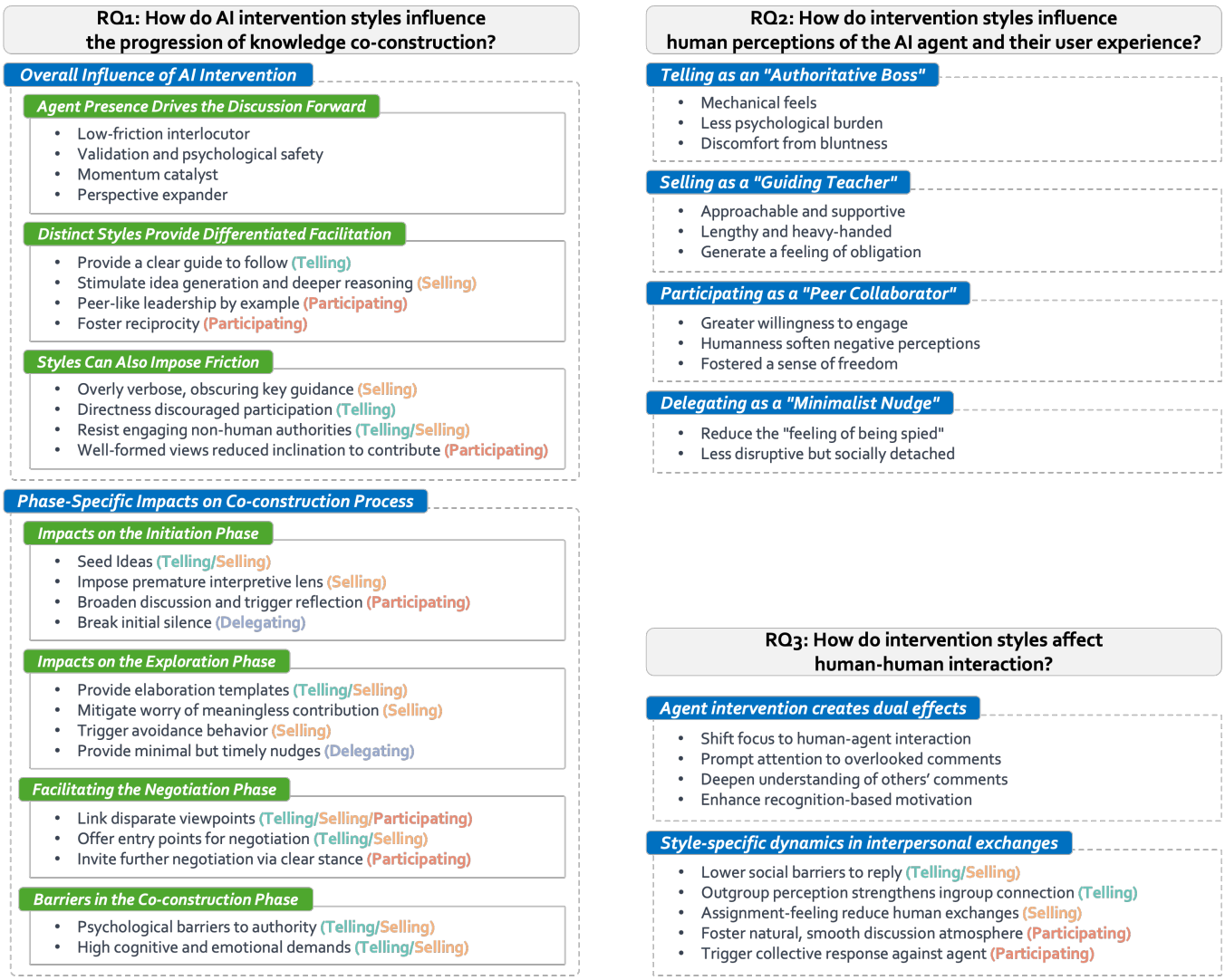


Figure 11: Summary of the main results from the thematic analysis, organized by research questions (RQs) and reported alongside quantitative findings in the corresponding Results subsections.

exploration phase; however, without an process-oriented orchestration, continued use of such cues after arguments are well-formed may restrain the natural progression toward integrative phases. Our experiment verifies that such process-orchestrated design can deepen knowledge co-construction in asynchronous settings. Participants recognized the agent's intent to advance the discussion, even through Delegating interventions, which often acted as *"timely situational cues"* (P16). Several noted that phase transitions felt natural because preceding phases were sufficiently developed (P23, P57), enabling smooth, organic progression of discussion threads. This orchestration logic is especially pertinent in asynchronous contexts. Unlike synchronous discussions, where shared temporal context can naturally drive progression [47], asynchronous platforms allow participants to join a thread at various stages, which may be underdeveloped, requiring elaboration, or already well-formed and ready for integration. At the collective level, such asynchronous entry points can lead to misaligned focus across participants—for example, late joiners to an advanced thread may revisit foundational issues, while long-time contributors are already synthesizing ideas. This divergence can fragment the discussion into parallel, uncoordinated sub-conversations, reducing the efficiency of collective knowledge advancement. By making phase readiness and transitions explicit, the agent maintains shared orientation, supporting steadier progression toward deeper co-construction.

However, our implementation revealed a limitation: mismatches between the agent's sufficiency judgments and participants' perceived readiness. Sometimes when discussions stagnated in Phase 2, the agent triggered a Phase-2-to-Phase-3 transition, some participants experienced it as interruptive because they had not yet *"talked*

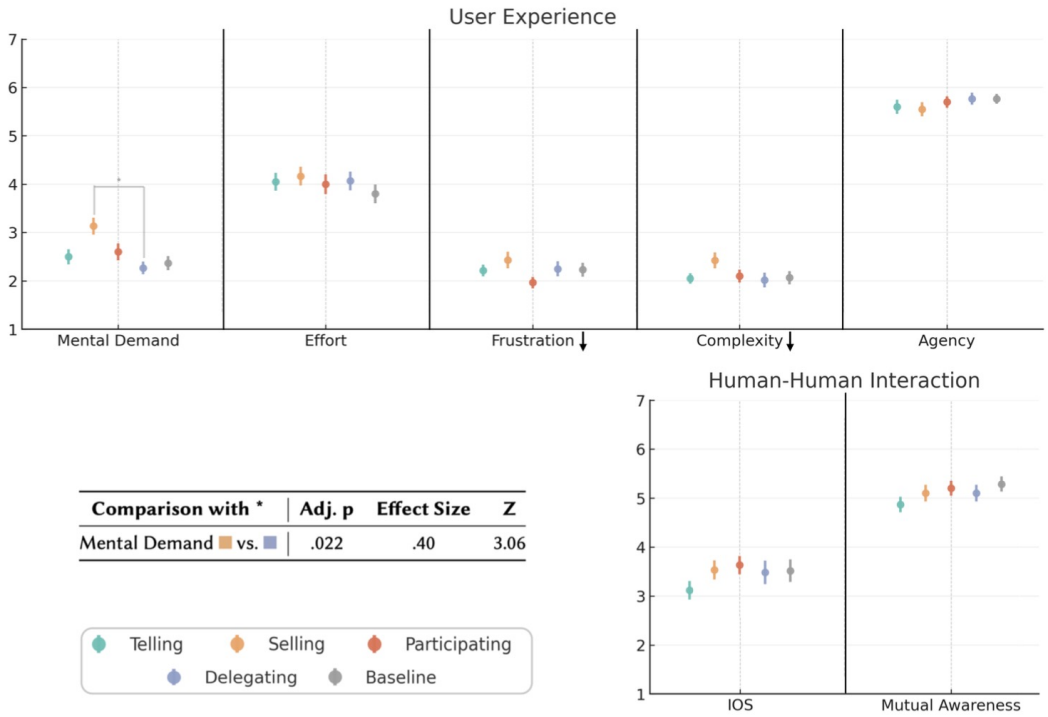


Figure 12: Participants' perceptions of (a) their user experience and (b) human-human interactions in the discussion across different conditions. Error bars represent 95% confidence intervals (*$p < .050$, **$p < .010$, ***$p < .001$). For significant pairwise comparisons, test statistics and effect sizes are reported. Unless marked with ↓ (indicating lower is better), higher scores denote better performance.

it through". To mitigate such dissonance, future designs could **humanize sufficiency assessment by incorporating social readiness signals**—such as participation dispersion, recent uptake of new ideas, or decreasing redundancy—alongside content-based indicators. Such adjustments may preserve perceived autonomy and reduce psychological reactance [18, 31], while still enabling timely, phase-appropriate progression.

Although developed for knowledge co-construction, the process-orchestrated logic can extend to other multi-phase collaborative tasks where later stages presuppose the successful completion of earlier ones. Examples include collaborative writing [87] and project-based online learning [15]. However, tasks without sequential dependencies—such as parallel idea listing or casual social chatting—are less suitable. In these cases, non-linear exploration is a core affordance [111], and imposing staged progression may constrain creativity and disrupt flow [10, 25]. Similarly, in time-critical collaboration (e.g., emergency response coordination [126]), our intervention logic may be counterproductive, as speed of convergence outweighs the value of building exhaustive grounding.

6.2 Navigating the Tension between Task-Oriented and Relationship-Oriented Intervention Styles

Telling, Selling, and Participating produced comparable objective outcomes (Table 3), likely because each style carries distinct pros and cons. Telling (high-task, low-relationship) provide clear and explicit guidance. Yet its authoritative delivery often made participants feel *"managed by a bot"* (P56) or *"interviewed"* (P17), prompting resistance. Participating (low-task, high-relationship) preserved equality and peer-like rapport, fostering willingness to respond. However, its subtle guidance sometimes failed to provide needed directional input, resulting in stagnation. This trade-off parallels findings in group leadership research: high-directive styles can compromise perceived autonomy and intrinsic motivation [30, 83], while purely relationship-oriented approaches risk under-scaffolding, leading to reduced coordination and goal alignment [64, 139].

Selling (high-task, high-relationship) has a great potential to reconcile this tension - preserving accessibility while providing structural guidance. Literature suggests that self-explanatory directives can reduce hierarchy perception and strengthen legitimacy [49], and that suggestions, rather than prescriptions, can support autonomy while sustaining progress [33, 115]. However, our experiment revealed that Selling did not fully realize this promise. Its

"guiding teacher" tone was indeed perceived as more personable than Telling's *"authoritative boss"*, and its persuasive explanations made its instructions more acceptable. But as mentioned in Section 5.2.2, participants felt an implied obligation, which increased emotional burden and sometimes triggered avoidance. This echoes emotional coercion mechanisms where obligation is used to gain compliance [100], and when such tactics are perceived, people may exhibit heightened psychological reactance, reduced trust in the source, and engage in withdrawal or resistance to protect autonomy [18]— patterns evident in our observations. This backfire effect may also stem from participants' perceptions of AI's social roles. Even for Selling, the agent implicitly positioned itself as a superior to the human in the interaction, leading some participants (12/60) to ask, *"Why should an AI instruct me?"* This reflects a mismatch between the role AI adopted in the study and human expectations, which appeared to trigger instinctive resistance. As AI's roles in society continue to evolve, selling-style interventions may gain acceptance in contexts where such authority feels legitimate.

To address this tension, designers may draw on the principle of SLM, which posits that there is no "best" style, and optimal influence is achieved by aligning behaviors with task demands and follower readiness [63]. In our context, task demands vary across phases of knowledge co-construction, with differing information needs and engagement patterns. Therefore, a promising approach is to **adapt the AI agent's style dynamically to the evolving phase of knowledge co-construction**, mitigating the limitations of any single style while leveraging its strengths when most advantageous. Our phase-specific insights (see Section 5.2.3) can serve as an empirical foundation for this strategy.

A potential trade-off in style adaptation is loss of a consistent agent persona. Frequent shifts in communicative style could challenge users' ability to form stable expectations about the agent, potentially reducing trust or perceived reliability [81, 131]. One complementary solution is a **multi-agent intervention framework**, in which specialized agents maintain distinct, internally coherent styles aligned with their facilitation role and intervene only in phases where their style is most effective. In some cases, this design can enable parallel and complementary interventions. For example, when a discussion needs to transition into the negotiation phase, a selling-style agent can prompt participants to initiate negotiation around those discrepancies, and a participating-style agent can respond by adopting a clear position to reconcile differing viewpoints. Operating in parallel, the agents might generate synergistic effects and better address the heterogeneous needs of users [16, 125]. Such multi-agent design can reduce users' cognitive load of interpreting fluctuating agent behaviors, allowing each agent to build rapport and credibility within its defined scope of responsibility [16, 22].

Regarding the follower readiness dimension of SLM, while participants are typically fluid in asynchronous platforms, it is still applicable in platforms that support private messaging. By drawing on interaction history to estimate a user's contribution ability, motivation, and receptivity to different styles, the agent could **deliver personalized interventions through direct messages.**

6.3 Other Design Considerations

6.3.1 Attention is a scarce resource: AI should serve as glue, not gravity. Participants reported two contrasting attentional effects of the agent on human–human interaction. Acting as *glue*, AI amplified the visibility and accessibility of overlooked contributions by elaborating on them, bridging knowledge gaps and fostering resonance among participants who might otherwise disengaged. As *gravity*, however, it drew interactions toward itself; participants replied to AI rather than to one another, particularly when it adopted a directive style. Excessively comprehensive responses further reduced opportunities for others to build upon contributions, weakening human-human ties.

To enhance human-human interaction, intervention design should follow a *one-intention-per-turn* heuristic: deliver a single move or intention at a time. If two intentions are necessary (e.g., seeding an idea and explaining its relevance), split them across turns with time in between. This reduces cognitive load and preserves space for peer contributions, aligning with best practices in facilitating distributed collaboration [33, 120]. Additionally, message frequency and length should be moderated to avoid monopolizing attention. Moreover, we suggest continuously monitoring the *interaction edge ratio* – the proportion of human–human edges relative to human–AI edges – as a live indicator of interaction balance. A rising dominance of human–AI edges signals a drift toward gravity, warranting shorter, less frequent interventions. Under a phase–style adaptive framework (Section 6.2), this metric can also inform style selection: when human–AI edges dominate, task-oriented styles should give way to more participatory approaches.

6.3.2 Recognition matters: even bot replies reduce the "void". A salient finding is that receiving any reply – human or AI – can legitimize participation. Several participants noted that even a brief acknowledgment from a bot reduced the sense of *"shouting into the void"* and of being ignored, thereby lowering the perceived social cost of contributing again (P3, P24). This effect aligns with the Computers Are Social Actors theory, which posits that humans tend to anthropomorphize responsive systems, applying interpersonal norms to their interactions [99]. Therefore, when an agent replies, it is often perceived as a socially meaningful response that affirms the contributor's presence. Such acknowledgment can activate reciprocity norms common in online communities, which in turn encourage re-engagement [80, 124].

Such recognition may be particularly valuable for novice contributors, who experience greater uncertainty and social risk when entering a discussion [80]. AI agents can prioritize detecting contributions from newcomers and selectively offering validation or praise, creating low-cost, high-impact interventions to foster continued participation. However, excessive or indiscriminate recognition should be avoided, as it may shift conversational gravity toward the AI (Section 6.3.1), diminish the perceived value of approval, weaken perceived sincerity, and potentially undermine intrinsic motivation [29, 59] .

6.3.3 Shifting synthesis: from manual production to assisted refinement. The result indicates that expecting users to independently synthesize knowledge in Phase 4 is often unrealistic, although workshop participants suggested that active synthesis and contributor

agency are critical for solidifying consensus (Section 3.2). We observed that directive interventions (Telling/Selling) largely failed because participants perceived summarization not merely as labor, but as an act of claiming authority over the group's knowledge. Consequently, users are inclined to relinquish agency in this phase when encountering these cognitive and social barriers. However, fully automating this phase risks reducing contributors to passive consumers, bypassing the cognitive engagement required to consolidate a shared understanding [44]. We therefore propose a design shift from *human production* to *human refinement and reflection*. Instead of asking users to synthesize from scratch, agents should act as "draftsmen," generating an initial summary. This approach lowers cognitive barriers and mitigates social pressure [60, 61]: users are no longer "claiming authority" over peers but simply refining and reflecting upon an AI's artifact. Thus, user agency is preserved through the higher-order tasks of validation and reflection—ensuring the consensus remains human-verified while offloading the labor of formulation.

6.3.4 Ethical Concerns.

Responsible Use of LLMs in Knowledge Communication: LLMs are advancing rapidly but remain prone to "hallucinations," producing plausible yet inaccurate information [70]. In one-to-one knowledge exchange, such misinformation can be highly detrimental; however, in knowledge co-construction settings, its impact may be less severe. Online communities often exhibit robust self-regulation mechanism, where participants collectively identify and correct inaccuracies, thus mitigating the spread of misinformation [138, 142]. In some cases, such collaborative misinformation correction can deepen discourse from exploration toward negotiation [57]. Nonetheless, when designing LLM-based agents as active community members, researchers should adopt a cautious and transparent approach, making users aware of the potential for error and the inherent limitations of LLM-generated content.

Minimizing Intrusiveness in Personal Probing: As discussed in Section 5.3.2, when agents ask increasingly specific questions about sensitive, identifiable information—such as participants' real-life occupations—users may perceive the interaction as overly intrusive, leading to discomfort and anxiety. To safeguard trust and address privacy concerns, researchers should design interaction protocols that limit the depth of personal probing and ensure transparency about why such information is requested and how it contributes to the collective knowledge-building process.

Preserving Human Role in AI-Supported Knowledge Co-Construction: During the interview, P32 stated that the participating-style agent appeared *"too intelligent,"* expressing uncertainty about whether such agents might eventually replace all human members, leaving AI entities to converse exclusively with one another in the co-construction of knowledge. This highlights an ethical concern: although AI agents can enhance the depth of discussions, they may also lead participants to fear that their own contributions will be marginalized, thereby reducing their motivation to engage and undermining the social and educational value of active human involvement. To sustain human motivation and perceived contribution, designs could incorporate transparent source attribution for agent-generated knowledge, especially when it derives from public, human-authored materials. By explicitly acknowledging these origins, the system can convey that its outputs are grounded in existing human knowledge rather than originating solely from the AI itself. Additional approaches may involve deliberately leaving certain knowledge gaps to prompt human participants to contribute, thereby sustaining their active role in the co-construction process.

6.4 Limitation and Future Work

This study has several limitations. **First**, our within-subject laboratory design—where each participant engaged in one discussion thread per day within a fixed six-person group over five consecutive days—enabled tight experimental control, minimized data contamination, and partially simulated authentic knowledge co-construction processes. However, this setting does not fully capture real-world dynamics, where participants of a thread are fluid throughout the day, and individuals often join multiple threads per day. Moreover, while our threads resembled small- to medium-scale discussions on real platforms, large-scale threads with thousands of comments may exhibit different interaction patterns and knowledge-building dynamics. Therefore, while the present findings provide valuable evidence, future research should conduct larger-scale, longer-term field studies in fully naturalistic contexts. **Second**, the participant sample consisted predominantly of young adults (ages 18–31) from East Asian cultural backgrounds. Further investigations should extend to participants of more diverse ages and cultural backgrounds. **Third**, to ensure that all participants could meaningfully contribute, we selected accessible, everyday discussion topics. While appropriate for our recruitment scope, such topics may elicit less "hardcore" or technically specialized knowledge exchange compared to scientific topics. Future research should therefore investigate the effectiveness of agent interventions in more knowledge-intensive subject contexts. **Fourth**, our interface adopted a chronological layout; alternative designs, such as tree-structured layouts, may shape discussion dynamics differently. Future research should compare agent effects across platforms with varied front-end interfaces. **Fifth**, we acknowledge the potential for a demand effect [98, 105], where participants feel compelled to "play along" with the agent. Our study did not measure this effect directly; future work should evaluate it in more naturalistic contexts where users have full autonomy to ignore interventions, thereby distinguishing such compliance from users' genuine preferences.

7 Conclusion

In this paper, we propose a process-orchestrated intervention paradigm for AI agents that progressively advance knowledge co-construction in asynchronous online discussions. The specific intervention strategies for task-oriented and relationship-oriented styles within this paradigm were derived from a design workshop. Implementing these strategies in an LLM-based agent, we conducted a within-subject study to examine their potential impact on knowledge co-construction dynamics, user perceptions and experiences, and human–human interactions. Our results demonstrate the promise of such intervention logic in fostering deeper and more robust collective knowledge. They also reveal the differentiated effects of task- and relationship-oriented styles on the co-construction process and participant behaviors, leading to several design considerations and research directions. With these empirical insights,

we position this work as an exploratory step toward developing contextual AI agents that more effectively enhance online discourse as an active member.

Acknowledgments

This project is supported by the Hong Kong SAR Research Grants Council's Theme-based Research Grant Scheme (Project No. T43-518/24-N). We also extend our gratitude to the reviewers for their constructive feedback, which had significantly improved the quality of our paper.

References

[1] [n. d.]. VooV Meeting. https://voovmeeting.com/. Accessed: 2025-09-09.
[2] 2025. 1% rule (Internet culture). https://en.wikipedia.org/wiki/1%25_rule. Accessed: 2025-09-10.
[3] Nimer Alrushiedat and Lorne Olfman. 2012. Anchored asynchronous online discussions: Facilitating participation and engagement in a blended environment. In *2012 45th Hawaii International Conference on System Sciences*. IEEE, 11–20.
[4] Arthur Aron, Elaine N Aron, and Danny Smollan. 1992. Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of personality and social psychology* 63, 4 (1992), 596.
[5] ATLAS.ti Scientific Software Development GmbH. 2023. ATLAS.ti Mac (version 23.2.1) [Qualitative data analysis software]. https://atlasti.com. Accessed: 2025-11-23.
[6] Tal August, Dallas Card, Gary Hsieh, Noah A. Smith, and Katharina Reinecke. 2020. Explain like I am a Scientist: The Linguistic Barriers of Entry to r/science. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–12. doi:10.1145/3313831.3376524
[7] Sandeep Avula, Bogeum Choi, and Jaime Arguello. 2022. The effects of system initiative during conversational collaborative search. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–30.
[8] Aadesh Bagmar, Kevin Hogan, Dalia Shalaby, and James Purtilo. 2022. Analyzing the effectiveness of an extensible virtual moderator. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–16.
[9] Roy F Baumeister and Mark R Leary. 1997. Writing narrative literature reviews. *Review of general psychology* 1, 3 (1997), 311–320.
[10] Benjamin B Bederson. 2004. Interfaces for staying in the flow. *Ubiquity* 5, 27 (2004), 1.
[11] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 57, 1 (1995), 289–300.
[12] Ivo Benke, Michael Thomas Knierim, and Alexander Maedche. 2020. Chatbot-based emotion management for distributed teams: A participatory design study. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–30.
[13] Alison Black. 2005. The use of asynchronous discussion: Creating a text of talk. *Contemporary issues in Technology and teacher education* 5, 1 (2005), 5–24.
[14] Robert Blake and Jane Mouton. 1964. The managerial grid: The key to leadership excellence. *Houston: Gulf Publishing Co* 350 (1964).
[15] Phyllis C Blumenfeld, Elliot Soloway, Ronald W Marx, Joseph S Krajcik, Mark Guzdial, and Annemarie Palincsar. 1991. Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational psychologist* 26, 3-4 (1991), 369–398.
[16] Nathan Bos, Judy Olson, Darren Gergle, Gary Olson, and Zach Wright. 2002. Effects of four computer-mediated communications channels on trust development. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 135–140.
[17] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
[18] Jack W Brehm. 1966. A theory of psychological reactance. (1966).
[19] Norman E Breslow and David G Clayton. 1993. Approximate inference in generalized linear mixed models. *Journal of the American statistical Association* 88, 421 (1993), 9–25.
[20] Zana Buçinca, Phoebe Lin, Krzysztof Z Gajos, and Elena L Glassman. 2020. Proxy tasks and subjective measures can be misleading in evaluating explainable AI systems. In *Proceedings of the 25th international conference on intelligent user interfaces*. 454–464.
[21] Zhenyao Cai, Seehee Park, Nia Nixon, and Shayan Doroudi. 2024. Advancing Knowledge Together: Integrating Large Language Model-based Conversational AI in Small Group Collaborative Learning. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–9. doi:10.1145/3613905.3650868
[22] Ana Paula Chaves and Marco Aurelio Gerosa. 2021. How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction* 37, 8 (2021), 729–758.
[23] XinHui Chen, Xiang Yuan, Hui Zhang, Ruixiao Zheng, and Wanyi Wei. 2025. Maintaining" Balanced" Conflict: Proactive Intervention Strategies of AI Voice Agents in Online Collaboration of Temporary Design Teams. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19.
[24] 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*. 103–119.
[25] Mihaly Csikszentmihalyi et al. 1997. Flow and the psychology of discovery and invention. *HarperPerennial, New York* 39 (1997), 1–16.
[26] Sander de Jong, Joel Wester, Tim Schrills, Kristina S. Secher, Carla F. Griggio, and Niels van Berkel. 2024. Assessing Cognitive and Social Awareness among Group Members in AI-assisted Collaboration. In *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia*. 338–350.
[27] Dhanielly PR de Lima, Marco A Gerosa, Tayana U Conte, and José Francisco de M. Netto. 2019. What to expect, and how to improve online discussion forums: the instructors' perspective. *Journal of Internet Services and Applications* 10, 1 (2019), 22.
[28] Bram De Wever, Tammy Schellens, Martin Valcke, and Hilde Van Keer. 2006. Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & education* 46, 1 (2006), 6–28.
[29] Edward L Deci, Richard Koestner, and Richard M Ryan. 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin* 125, 6 (1999), 627.
[30] Edward L Deci and Richard M Ryan. 1987. The support of autonomy and the control of behavior. *Journal of personality and social psychology* 53, 6 (1987), 1024.
[31] Edward L Deci and Richard M Ryan. 2000. The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry* 11, 4 (2000), 227–268.
[32] Ana Del Pino-Marchito, Agustín Galán-García, and María de los Ángeles Plaza-Mejía. 2025. The Hersey and Blanchard's Situational Leadership Model Revisited: Its Role in Sustainable Organizational Development. *World* 6, 2 (2025), 1–28.
[33] Pierre Dillenbourg. 2002. Over-scripting CSCL: The risks of blending collaborative learning with instructional design. *Three worlds of CSCL. Can we support CSCL?* (2002), 61–91.
[34] Discourse Contributors. 2014. Discourse – Civilized Discussion for Your Community. https://www.discourse.org/. Accessed: 2025-09-10.
[35] 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 (2022), 1–23.
[36] Hyo Jin Do, Ha-Kyung Kong, Pooja Tetali, Karrie Karahalios, and Brian P Bailey. 2023. Inform, Explain, or Control: Techniques to Adjust End-User Performance Expectations for a Conversational Agent Facilitating Group Chat Discussions. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–26.
[37] Emily Doherty, E Margaret Perkoff, Sean von Bayern, Rui Zhang, Indrani Dey, Michal Bodzianowski, Sadhana Puntambekar, and Leanne Hirshfield. 2025. Piecing together teamwork: A responsible approach to an LLM-based educational jigsaw agent. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–17.
[38] Zhao Du, Fang Wang, Shan Wang, and Xiao Xiao. 2022. Enhancing learner participation in online discussion forums in massive open online courses: the role of mandatory participation. *Frontiers in psychology* 13 (2022), 819640.
[39] Ilana Dubovi and Iris Tabak. 2020. An empirical analysis of knowledge co-construction in YouTube comments. *Computers & Education* 156 (Oct. 2020), 103939. doi:10.1016/j.compedu.2020.103939
[40] Thomas Erickson and Susan C Herring. 2005. Persistent conversation: A dialog between research and design. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, Vol. 5. IEEE Computer Society, 106–106.
[41] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191.
[42] Fred E Fiedler. 1967. A THEORY OF LEADERSHIP EFFECTIVENESS. MCGRAW-HILL SERIES IN MANAGEMENT. (1967).
[43] Frank Fischer, Ingo Kollar, Karsten Stegmann, and Christof Wecker. 2013. Toward a script theory of guidance in computer-supported collaborative learning. *Educational psychologist* 48, 1 (2013), 56–66.
[44] James Fishkin. 2009. *When the people speak: Deliberative democracy and public consultation*. Oup Oxford.
[45] Donelson R Forsyth. 2014. *Group dynamics*. Wadsworth Cengage Learning.
[46] Milton Friedman. 1940. A comparison of alternative tests of significance for the problem of m rankings. *The annals of mathematical statistics* 11, 1 (1940), 86–92.

[47] Hugo Fuks, Alberto B Raposo, Marco A Gerosa, and Carlos JP Lucena. 2005. Applying the 3C model to groupware development. *International Journal of Cooperative Information Systems* 14, 02n03 (2005), 299–328.

[48] Jarod Govers, Eduardo Velloso, Vassilis Kostakos, and Jorge Goncalves. 2024. AI-Driven Mediation Strategies for Audience Depolarisation in Online Debates. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.

[49] Bernard Guerin. 1994. *Analyzing social behavior: Behavior analysis and the social sciences*. Context Press.

[50] Charlotte Nirmalani Gunawardena, Nick V Flor, David Gomez, and Damien Sanchez. [n. d.]. Analyzing Social Construction of Knowledge Online by Employing Interaction Analysis, Learning Analytics, and Social Network Analysis. ([n. d.]).

[51] Charlotte N. Gunawardena, Constance A. Lowe, and Terry Anderson. 1997. Analysis of a Global Online Debate and the Development of an Interaction Analysis Model for Examining Social Construction of Knowledge in Computer Conferencing. *Journal of Educational Computing Research* 17, 4 (Dec. 1997), 397–431. doi:10.2190/7MQV-X9UJ-C7Q3-NRAG

[52] Rafik Hadfi, Jawad Haqbeen, Sofia Sahab, and Takayuki Ito. 2021. Argumentative Conversational Agents for Online Discussions. *Journal of Systems Science and Systems Engineering* 30, 4 (Aug. 2021), 450–464. doi:10.1007/s11518-021-5497-1 Publisher: Springer Science and Business Media LLC.

[53] Reza Hadi Mogavi, Bingcan Guo, Yuanhao Zhang, Ehsan-Ul Haq, Pan Hui, and Xiaojuan Ma. 2022. When gamification spoils your learning: A qualitative case study of gamification misuse in a language-learning app. In *Proceedings of the ninth ACM conference on learning@ scale*. 175–188.

[54] Reza Hadi Mogavi, Yuanhao Zhang, Ehsan-Ul Haq, Yongjin Wu, Pan Hui, and Xiaojuan Ma. 2022. What do users think of promotional gamification schemes? a qualitative case study in a question answering website. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–34.

[55] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[56] Changyang He, Yue Deng, Lu He, Qingyu Guo, Yu Zhang, Zhicong Lu, and Bo Li. 2024. Engage Wider Audience or Facilitate Quality Answers? a Mixed-methods Analysis of Questioning Strategies for Research Sensemaking on a Community Q&A Site. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (April 2024), 1–31. doi:10.1145/3637327

[57] Changyang He, Lu He, Tun Lu, and Bo Li. 2021. Beyond entertainment: Unpacking Danmaku and comments' role of information sharing and sentiment expression in online crisis videos. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–27.

[58] Selina Heller, Mohamed Ibrahim, David Antony Selby, and Sebastian Vollmer. 2025. Finding Common Ground: Using Large Language Models to Detect Agreement in Multi-Agent Decision Conferences. *arXiv preprint arXiv:2507.08440* (2025).

[59] Jennifer Henderlong and Mark R Lepper. 2002. The effects of praise on children's intrinsic motivation: a review and synthesis. *Psychological bulletin* 128, 5 (2002), 774.

[60] John Heritage. 2012. Epistemics in conversation. *The handbook of conversation analysis* (2012), 370–394.

[61] John Heritage and Geoffrey Raymond. 2005. The terms of agreement: Indexing epistemic authority and subordination in talk-in-interaction. *Social psychology quarterly* 68, 1 (2005), 15–38.

[62] Susan C Herring. 1999. Interactional coherence in CMC. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*. IEEE, 13–pp.

[63] Paul Hersey, Kenneth H Blanchard, and Walter E Natemeyer. 1979. Situational leadership, perception, and the impact of power. *Group & organization studies* 4, 4 (1979), 418–428.

[64] Robert J House. 1996. Path-goal theory of leadership: Lessons, legacy, and a reformulated theory. *The leadership quarterly* 7, 3 (1996), 323–352.

[65] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685* (2021).

[66] Zeyu Huang, Xinyi Cao, Yuanhao Zhang, and Xiaojuan Ma. 2024. Sharing Frissons among Online Video Viewers: Exploring the Design of Affective Communication for Aesthetic Chills. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19.

[67] Angel Hsing-Chi Hwang and Andrea Stevenson Won. 2022. Ai in your mind: Counterbalancing perceived agency and experience in human-ai interaction. In *Chi conference on human factors in computing systems extended abstracts*. 1–10.

[68] Takayuki Ito, Rafik Hadfi, and Shota Suzuki. 2022. An Agent that Facilitates Crowd Discussion: A Crowd Discussion Support System based on an Automated Facilitation Agent. *Group Decision and Negotiation* 31, 3 (June 2022), 621–647. doi:10.1007/s10726-021-09765-8

[69] Takayuki Ito, Rafik Hadfi, and Shota Suzuki. 2022. An agent that facilitates crowd discussion: A crowd discussion support system based on an automated facilitation agent. *Group Decision and Negotiation* 31, 3 (2022), 621–647.

[70] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *ACM computing surveys* 55, 12 (2023), 1–38.

[71] Albert Jiang, Edouard Grave, Chloé Bamford, et al. 2023. Mistral 7B. https://mistral.ai/news/introducing-mistral-7b/. Accessed: 2025-09-10.

[72] Ridley Jones, Lucas Colusso, Katharina Reinecke, and Gary Hsieh. 2019. r/science: Challenges and Opportunities in Online Science Communication. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–14. doi:10.1145/3290605.3300383

[73] Celia Kaendler, Michael Wiedmann, Nikol Rummel, and Hans Spada. 2015. Teacher competencies for the implementation of collaborative learning in the classroom: A framework and research review. *Educational Psychology Review* 27, 3 (2015), 505–536.

[74] Hanseob Kim, Bin Han, Jieun Kim, Muhammad Firdaus Syawaludin Lubis, Gerard Jounghyun Kim, and Jae-In Hwang. 2024. Engaged and affective virtual agents: Their impact on social presence, trustworthiness, and decision-making in the group discussion. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.

[75] 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*. 1–13.

[76] Soomin Kim, Jinsu Eun, Joseph Seering, and Joonhwan Lee. 2021. Moderator chatbot for deliberative discussion: Effects of discussion structure and discussant facilitation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–26.

[77] Joachim Kimmerle, Johannes Moskaliuk, Aileen Oeberst, and Ulrike Cress. 2015. Learning and Collective Knowledge Construction With Social Media: A Process-Oriented Perspective. *Educational Psychologist* 50, 2 (April 2015), 120–137. doi:10.1080/00461520.2015.1036273

[78] Alison King. 1997. ASK to THINK-TEL WHY: A model of transactive peer tutoring for scaffolding higher level complex learning. *Educational psychologist* 32, 4 (1997), 221–235.

[79] Ingo Kollar, Frank Fischer, and Friedrich W Hesse. 2006. Collaboration scripts–a conceptual analysis. *Educational Psychology Review* 18, 2 (2006), 159–185.

[80] Cliff Lampe and Paul Resnick. 2004. Slash (dot) and burn: distributed moderation in a large online conversation space. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 543–550.

[81] Kwan Min Lee, Namkee Park, and Hayeon Song. 2005. Can a robot be perceived as a developing creature? Effects of a robot's long-term cognitive developments on its social presence and people's social responses toward it. *Human communication research* 31, 4 (2005), 538–563.

[82] Jennifer Lee-Baldwin. 2005. Asynchronous discussion forums: A closer look at the structure, focus and group dynamics that facilitate reflective thinking. *Contemporary Issues in Technology and Teacher Education* 5, 1 (2005), 93–115.

[83] Kurt Lewin, Ronald Lippitt, and Ralph K White. 1939. Patterns of aggressive behavior in experimentally created "social climates". *The Journal of social psychology* 10, 2 (1939), 269–299.

[84] Jiansheng Li, Linlin Li, Zhixin Zhu, and Rustam Shadiev. 2023. Research on the predictive model based on the depth of problem-solving discussion in MOOC forum. *Education and Information Technologies* 28, 10 (2023), 13053–13076.

[85] Chengzhong Liu, Shixu Zhou, Dingdong Liu, Junze Li, Zeyu Huang, and Xiaojuan Ma. 2023. CoArgue: Fostering Lurkers' Contribution to Collective Arguments in Community-based QA Platforms. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[86] Chengzhong Liu, Shixu Zhou, Yuanhao Zhang, Dingdong Liu, Zhenhui Peng, and Xiaojuan Ma. 2022. Exploring the effects of self-mockery to improve task-oriented chatbot's social intelligence. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*. 1315–1329.

[87] Paul Benjamin Lowry, Aaron Curtis, and Michelle René Lowry. 2004. Building a taxonomy and nomenclature of collaborative writing to improve interdisciplinary research and practice. *The Journal of Business Communication (1973)* 41, 1 (2004), 66–99.

[88] Li-Fen Lilly Lu and Ifeng Jeng. 2006. Knowledge Construction in Inservice Teacher Online Discourse: Impacts of Instructor Roles and Facilitative Strategies. (2006).

[89] Margarida Lucas, Charlotte Gunawardena, and António Moreira. 2014. Assessing social construction of knowledge online: A critique of the interaction analysis model. *Computers in Human Behavior* 30 (Jan. 2014), 574–582. doi:10.1016/j.chb.2013.07.050

[90] Shuai Ma, Qiaoyi Chen, Xinru Wang, Chengbo Zheng, Zhenhui Peng, Ming Yin, and Xiaojuan Ma. 2025. Towards human-ai deliberation: Design and evaluation of llm-empowered deliberative ai for ai-assisted decision-making. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–23.

[91] Shuai Ma, Ying Lei, Xinru Wang, Chengbo Zheng, Chuhan Shi, Ming Yin, and Xiaojuan Ma. 2023. Who should i trust: Ai or myself? leveraging human and

ai correctness likelihood to promote appropriate trust in ai-assisted decision-making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[92] Shuai Ma, Mingfei Sun, and Xiaojuan Ma. 2023. Modeling adaptive expression of robot learning engagement and exploring its effects on human teachers. *ACM Transactions on Computer-Human Interaction* 30, 5 (2023), 1–48.

[93] Shuai Ma, Junling Wang, Yuanhao Zhang, Xiaojuan Ma, and April Yi Wang. 2025. Dbox: Scaffolding algorithmic programming learning through learner-llm co-decomposition. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.

[94] Shuai Ma, Xinru Wang, Ying Lei, Chuhan Shi, Ming Yin, and Xiaojuan Ma. 2024. "Are you really sure?" Understanding the effects of human self-confidence calibration in AI-assisted decision making. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–20.

[95] Shuai Ma, Taichang Zhou, Fei Nie, and Xiaojuan Ma. 2022. Glancee: An adaptable system for instructors to grasp student learning status in synchronous online classes. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–25.

[96] Catherine McLoughlin and Joe Luca. 2000. Cognitive engagement and higher order thinking through computer conferencing: We know why but do we know how. In *Flexible futures in tertiary teaching. Proceedings of the 9th Annual Teaching Learning Forum*, Vol. 24.

[97] Miro. [n. d.]. Miro Online Collaborative Whiteboard Platform. https://miro.com/. Accessed: 2025-09-09.

[98] Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of social issues* 56, 1 (2000), 81–103.

[99] 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*. 72–78.

[100] Cory F Newman. 1997. Emotional Blackmail: When the People in Your Life Use Fear, Obligation, and Guilt to Manipulate You. *Journal of Cognitive Psychotherapy* 11, 4 (1997), 300.

[101] Ha Nguyen. 2023. Role design considerations of conversational agents to facilitate discussion and systems thinking. *Computers & Education* 192 (2023), 104661.

[102] Ha Nguyen and Morgan Diederich. 2023. Facilitating knowledge construction in informal learning: A study of TikTok scientific, educational videos. *Computers & Education* 205 (Nov. 2023), 104896. doi:10.1016/j.compedu.2023.104896

[103] Blair Nonnecke and Jenny Preece. 2001. Why lurkers lurk. (2001).

[104] Javier Onrubia and Anna Engel. 2009. Strategies for collaborative writing and phases of knowledge construction in CSCL environments. *Computers & Education* 53, 4 (Dec. 2009), 1256–1265. doi:10.1016/j.compedu.2009.06.008

[105] Martin T Orne. 2017. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. In *Sociological methods*. Routledge, 279–299.

[106] Roy D Pea. 2018. The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. In *Scaffolding*. Psychology Press, 423–451.

[107] Zhenhui Peng, Taewook Kim, and Xiaojuan Ma. 2019. GremoBot: exploring emotion regulation in group chat. In *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing*. 335–340.

[108] Shaukat Ali Raza and Asma Sikandar. 2018. Impact of leadership style of teacher on the performance of students: An application of Hersey and Blanchard situational model. *Bulletin of Education and Research* 40, 3 (2018), 73–94.

[109] Reddit. 2005. https://www.reddit.com/

[110] Richard M Ryan and Edward L Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55, 1 (2000), 68.

[111] John Scott and Peter J Carrington. 2011. *The SAGE handbook of social network analysis*. SAGE publications.

[112] Samuel Sanford Shapiro and Martin B Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 3-4 (1965), 591–611.

[113] Donghoon Shin, Soomin Kim, Ruoxi Shang, Joonhwan Lee, and Gary Hsieh. 2023. IntroBot: Exploring the use of chatbot-assisted familiarization in online collaborative groups. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–13.

[114] Donghoon Shin, Sangwon Yoon, Soomin Kim, and Joonhwan Lee. 2021. Blahblahbot: Facilitating conversation between strangers using a chatbot with ml-infused personalized topic suggestion. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–6.

[115] Amy Soller. 2001. Supporting social interaction in an intelligent collaborative learning system. *International journal of artificial intelligence in education* 12 (2001), 40–62.

[116] StackExchange. 2019. https://stackexchange.com/

[117] Gerry Stahl. 2013. A model of collaborative knowledge-building. In *International conference of the learning sciences*. Psychology Press, 70–77.

[118] Candice R Stefanou, Kathleen C Perencevich, Matthew DiCintio, and Julianne C Turner. 2004. Supporting autonomy in the classroom: Ways teachers encourage student decision making and ownership. *Educational psychologist* 39, 2 (2004), 97–110.

[119] Ralph M Stogdill. 1974. *Handbook of leadership: A survey of theory and research*. Free Press.

[120] Daniel D Suthers. 2006. Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-supported collaborative learning* 1, 3 (2006), 315–337.

[121] Henri Tajfel, John Turner, William G Austin, and Stephen Worchel. 2001. An integrative theory of intergroup conflict. *Intergroup relations: Essential readings* (2001), 94–109.

[122] Seng Chee Tan, Alwyn Vwen Yen Lee, and Min Lee. 2022. A systematic review of artificial intelligence techniques for collaborative learning over the past two decades. *Computers and Education: Artificial Intelligence* 3 (2022), 100097. doi:10.1016/j.caeai.2022.100097

[123] Marijn Tanis, Marjolein Dobber, Rosanne Zwart, Bert van Oers, et al. 2014. *Beter leren door onderzoek: Hoe begeleid je onderzoekend leren van leerlingen?* Vrije Universiteit, Universiteit van Amsterdam.

[124] Lisa Collins Tidwell and Joseph B Walther. 2002. Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time. *Human communication research* 28, 3 (2002), 317–348.

[125] David Traum, Stacy C Marsella, Jonathan Gratch, Jina Lee, and Arno Hartholt. 2008. Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents. In *International workshop on intelligent virtual agents*. Springer, 117–130.

[126] Murray Turoff, Michael Chumer, Bartel Van de Walle, and Xiang Yao. 2004. The design of a dynamic emergency response management information system (DERMIS). *Journal of Information Technology Theory and Application (JITTA)* 5, 4 (2004), 3.

[127] Frans H Van Eemeren, A Francisca Sn Henkemans, and Rob Grootendorst. 2002. *Argumentation: Analysis, evaluation, presentation*. Routledge.

[128] Anouschka Van Leeuwen and Jeroen Janssen. 2019. A systematic review of teacher guidance during collaborative learning in primary and secondary education. *Educational Research Review* 27 (2019), 71–89.

[129] 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. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–25.

[130] Yunlong Wang, Shuyuan Shen, and Brian Y Lim. 2023. Reprompt: Automatic prompt editing to refine ai-generative art towards precise expressions. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–29.

[131] Adam Waytz, Joy Heafner, and Nicholas Epley. 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of experimental social psychology* 52 (2014), 113–117.

[132] Noreen M Webb. 2009. The teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology* 79, 1 (2009), 1–28.

[133] Noreen M Webb, Megan L Franke, Marsha Ing, Angela Chan, Tondra De, Deanna Freund, and Dan Battey. 2008. The role of teacher instructional practices in student collaboration. *Contemporary educational psychology* 33, 3 (2008), 360–381.

[134] Armin Weinberger and Frank Fischer. 2006. A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education* 46, 1 (Jan. 2006), 71–95. doi:10.1016/j.compedu.2005.04.003

[135] Evan James Williams. 1949. Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Chemistry* 2, 2 (1949), 149–168.

[136] Qianli Xu, Liyuan Li, and Gang Wang. 2013. Designing engagement-aware agents for multiparty conversations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2233–2242.

[137] Yogyakarta State University and Tri Wahyuni Floriasti. 2023. Analysis of Online Discussion Forum in Constructing English Language Skills Through Collaborative Learning of Preservice Teachers in Indonesia. *Mextesol Journal* 47, 1 (Jan. 2023), 1–12. doi:10.61871/mj.v47n1-4 Publisher: Asociación Mexicana de Maestros de Inglés MEXTESOL A.C..

[138] Kangyu Yuan, Li Zhang, Hanfang Lyu, Ziqi Pan, Yuanhao Zhang, Junze Li, Bingcan Guo, Jiaxiong Hu, Qingyu Guo, and Xiaojuan Ma. 2025. " I Love the Internet Again": Exploring the Interaction Inception of" TikTok Refugees" Flocking into RedNote. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.

[139] Gary Yukl. 2012. Leadership. *Cases in Leadership (3rd. ed. pp. 1–42). Thousand Oaks, CA: Sage* (2012).

[140] Rui Zhang, Nathan J McNeese, Guo Freeman, and Geoff Musick. 2021. " An ideal human" expectations of AI teammates in human-AI teaming. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–25.

[141] Yu Zhang, Changyang He, Huanchen Wang, and Zhicong Lu. 2023. Understanding Communication Strategies and Viewer Engagement with Science Knowledge Videos on Bilibili. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–18. doi:10.1145/3544548.3581476

[142] Yuanhao Zhang, Yumeng Wang, Xiyuan Wang, Changyang He, Chenliang Huang, and Xiaojuan Ma. 2025. CoKnowledge: Supporting Assimilation of Time-synced Collective Knowledge in Online Science Videos. *arXiv preprint arXiv:2502.03767* (2025).

[143] Chengbo Zheng, Yuanhao Zhang, Zeyu Huang, Chuhan Shi, Minrui Xu, and Xiaojuan Ma. 2024. DiscipLink: unfolding interdisciplinary information seeking process via human-AI co-exploration. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–20.

[144] Xinran Zhu, Hong Shui, and Bodong Chen. 2023. Beyond reading together: Facilitating knowledge construction through participation roles and social annotation in college classrooms. *The Internet and Higher Education* 59 (Oct. 2023), 100919. doi:10.1016/j.iheduc.2023.100919