"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

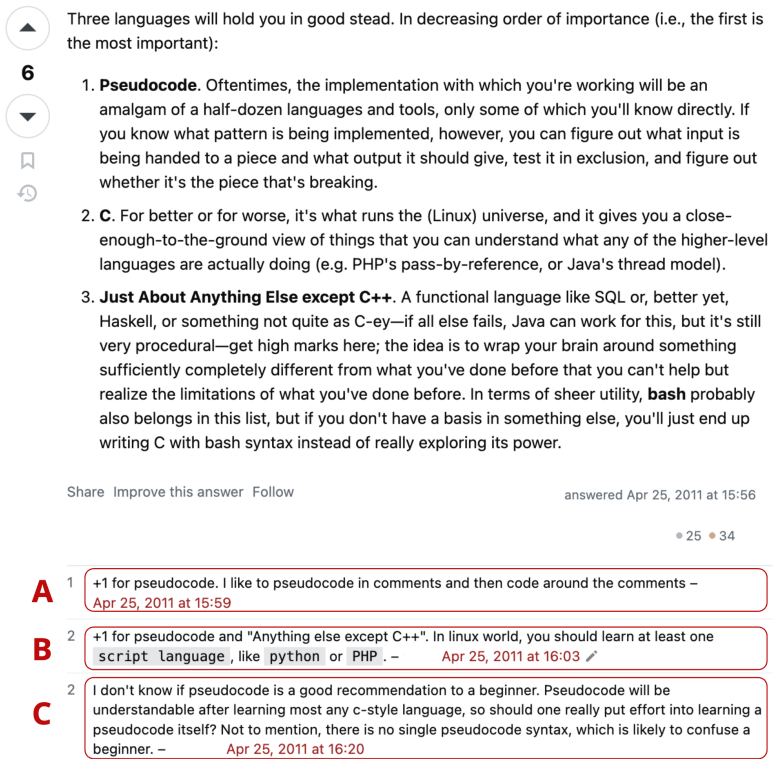

Figure 1: The screenshot shows a Stack Exchange discussion thread responding to the question "What language for starting on Linux?", where the accepted answer (top) recommends learning pseudocode, C, and any scripting language except C++. Comments A, B, and C co-construct knowledge by complementing, 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

speci c 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 signi cant 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 e ective for maintaining productive dynamics, these "coordinator" agents mainly manage the ow 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 rst 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 re ne and re ect 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-speci c 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 in uential models further conceptualized speci c styles within the space de ned by these two dimensions through a quadrant logic (e.g., Managerial Grid Model[14], Situational Leadership Model (SLM) [42, 63]).

TR framework also e ectively characterize the landscape of AI interventions for knowledge co-construction reviewed in Section 2.2. Speci cally, 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 de ne 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 di erent combinations of guidance (task) and peer-like participation (relationship) in uence 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. [103] reported greater group achievement from more task-oriented interventions. While these ndings concern human intervention, prior studies also suggest that participants are sensitive to the styles used by AI agents in group settings and may respond di erently 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, in uence 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 di erent styles. Insights from the workshop informed the development of an LLM-based agent.

3.1 Theoretical Framework

The design of strategies across di erent 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 re ned 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 identi ed in earlier discussions.
Phase4 Co-construction : Collaboratively integrating negotiation insights into a coherent synthesis, and re ecting 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-o 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 brie ng on the de nitions 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: De ning Su ciency Criteria for Each Phase (15 minutes) Building on theoretical de nitions, participants collaboratively speci ed the su ciency criteria for each phase. They shared perspectives, discussed di erences, 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 -Speci c 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 su ciency 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, prede ned phase-opening prompts without the need for extensive co-design.
Activity 3: Consolidating and Re ning Representative Strategies (75 minutes) All participants reconvened to exchange and re ne ideas generated in Activity 2, guided by a facilitator. New suggestions were welcomed during this stage. The process resulted in a nal 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 re ning both the su ciency 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 su ciency criteria for each phase and the representative intervention strategies for each agent style. Given the collaborative structure of the workshop, results re ect group consensus rather than individual frequencies: su ciency criteria were established by the full cohort ($\# = 12$), while strategies were synthesized in subgroups ($\# = 4$) and subsequently validated by all participants ($\# = 12$). 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, o ering 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 su ciency by examining the completeness of the argumentative structure. Telling-style agents could identify underexplored arguments and instruct participants to elaborate on them from speci c missing angles. Selling-style agents would encourage such elaboration by underscoring its potential

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

ID	Gender	Primary Platform	Freq. of Contribution	Con dence 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. Con dence level is self-reported on a 7-point Likert scale (1=Low, 7=High)

.

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

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

Negotiation: In this phase, su ciency 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 clari ed 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., bene ts, 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 nal phase focuses on collaboratively synthesizing agreements reached and linking them to practical applications or re ective insights. Participants identi ed the active process of synthesis rather than just the resulting artifact as the critical mechanism for solidifying consensus. Consequently, success was de ned as a collective synthesis where contributors explicitly articulate the agreed knowledge, supported by substantial re ection. In this phase, telling-style agents explicitly instruct participants to produce a holistic summary, specifying which issues to address and from which perspectives to re ect. Selling-style agents, in addition to providing guidance, advocate for the value of comprehensive summarizing and re ection. Participating-style agents, acknowledging the potential labor burden, initiate a partial summary with their own re ections to sca old 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

Table 2: Summary of ndings from the design workshop, including the desired outcomes for each knowledge co-construction phase and the corresponding intervention strategies in di erent 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 identi ed barriers to e ective 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 su cient skills to determine when and how to advance the discourse [48, 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 nal co-construction phase.

DR2 Phase Su ciency: 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 super cial consensus, producing fragile rather than durable knowledge [134]. The agent should therefore ensure that each phase meets the su ciency criteria de ned 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, o ering only the necessary input to sustain productive dialogue while leaving participants su cient 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 su ciency criteria de ned 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. Speci cally, 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 ve core components: Comment Analyzer, Frequency Controller, Phase-Su ciency 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 ow 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 Classi er. Consistent with previous empirical approaches [39, 102, 144], we treat each comment as a distinct analytical unit, categorized into one of ve categories corresponding to the knowledge co-construction phases (0 = no contribution to knowledge advancement, 1 4 represent Phases 1 4). To build this classi er, 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 o cial 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 re ned corpus of 12587 comments.

Two authors manually coded 1,500 comments randomly sampled from the dataset. They rst independently labeled an initial set of 100 comments to align interpretations of phase de nitions, then independently labeled the remainder. Disagreements were resolved through discussion; unresolved cases were adjudicated by a third coder. Using the 1500 labeled samples, we-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 classi er generalizes well across varied discussion topics.

2. Reply Identi er. 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 identi es the source comment. All comments and their identi ed 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 reactivate participation and prevent prolonged inactivity. 2. Comment threshold: if ten new comments have been posted without the current phase reaching su ciency, the agent intervenes; if su ciency 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-Su ciency Evaluator : This module determines whether the current phase has reached su ciency and, based on that assessment, selects the appropriate phase-specific strategy. In our controlled lab study, the workshop-derived su ciency criteria for Phases 1 4 were adapted to t the experimental context (Section 4):

Initiation : The goal of this phase is to elicit a diverse set of initial contributions. Therefore, su ciency 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 simpli ed three-item checklist presence of a Quali er, 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 su cient 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 su ciency criterion is that every identi ed con ict is addressed by corresponding Phase-3 comments and resolved to consensus. Speci cally, the agent categorizes consensus into three types beyond simple agreement: (1) Clari ed Disagreement (where divergence is clear but irreconcilable), (2) Conditional Agreement (where di erent positions hold under speci c 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-e ciency. 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 e ectiveness in detecting whether a con ict has been resolved to consensus.
Co-construction : Here, su ciency requires that Phase-4 comments collectively summarize all agreements and contain more than three metacognitive statements that either re ect on or apply the collective knowledge.

Style Manager: Each discussion thread in the experiment is assigned a prede ned 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-Su ciency 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 re ned through iterative prompt-engineering. The selected prompt is passed to the LLM to generate the nal 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 speci c 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 in uence. Participants could contribute new comments to the discussion via an input box at the bottom of the page or interact with speci c comments using "reply" and "like" buttons. To nudge engagement, a noti cation badge in the top-right navigation bar appeared whenever a participant received a reply. Clicking this noti cation 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 di erent 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 su cient interaction for meaningful discourse.

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 di erent 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 di erent topics. Each thread is discussed by a distinct group of 6 participants. No participant engaged in multiple threads simultaneously to avoid cross-thread confounding e ects. 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 ve 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 e ective way to learn a new language . A full list of topics is provided in the supplementary material.

4.2 Participants and Procedure

We rst conducted an a priori power analysis in G*Power [41] based on a repeated-measures ANOVA design with ve conditions. Assuming a medium e ect size (5 = 0”25), a signi cance threshold (U = 0”05), and a statistical power (1 V = 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 rst completed a pre-task survey to collect demographic information and con rm eligibility, followed by a task brie ng. They then participated in ve sequential one-day discussion threads, each involving a distinct topic. The topic for each day was xed 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 di culty (Section 4.1). To control for order and carryover e ects, we employed a balanced latin square design [135]. Given ve 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 di erent peers everyday, we generated six mutually disjoint sets of these blocks. Speci cally, 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 e ect. 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 nal 60 condition sequences one for each participant (Figure 5). This design ensures that: (1) every condition appeared equally often on each topic, e ectively neutralizing topic-speci c biases; and (2) no two participants followed the exact same trajectory, preventing xed subgroups.

Following each daily discussion, participants completed an in-task survey to provide condition-speci c feedback. Upon completing all threads, participants took part in a semi-structured interview to explore their knowledge co-construction experiences, perceptions of di erent 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; speci c 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 nal phase. Su ciency Criteria indicated whether the thread met prede ned requirements for each phase (Table 2). Two researchers independently coded these metrics based on the de nitions in Section 3, and resolved discrepancies through discussion. At the individual level, participants rated the discussion's Smoothness [37], E ectiveness [75, 76], Depth [89], and Su ciency [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], speci cally Mental Demand [70], E ort [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 a ect 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. Signi cant main e ects were followed by Bonferroni-corrected Wilcoxon signed-rank post-hoc tests. We report test statistics, adjusted p-values, and e ect 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 speci ed

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

as dependent variables, Condition as a xed e ect, and Day as a random intercept to control for day-speci c variability. We speci- ed model distributions based on variable type: binomial for binary outcomes (e.g., Phase Su ciency), negative binomial for count variables (e.g., Agent Comment Likes), and cumulative link models for ordinal variables (e.g., Max Phase Reached). Upon observing signi cant main e ects, 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 e ect 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 ve 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 [8, 54]. First, audio recordings were automatically transcribed via Zoom. Two researchers reviewed all transcripts to ensure delity 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 re ned through iterative team discussions. The nal themes and codes are summarized in Figure 11, which systematically maps speci c 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 o er 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 classi cations matched the intended style, suggesting that participants generally recognize the correct style. A Chi-square test con rmed that the observed classi cation frequencies signi cantly di ered from chance ($\chi^2_{19} = 432.58$, $p < .001$), indicating that the agents were reliably perceived in line with their designated styles. This veri ed that our manipulation of the independent variable was e ective, ensuring that our participants perceived, understood, and reacted to the four intended agent styles as expected.

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

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

Table 3: Descriptive statistics of thread-level measures across the ve 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 signi cance 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 signi cance 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-su ciency analysis reinforces this divergence. While 90% of baseline groups met su ciency in Phase 1, only 10% reached su ciency in Phase 2. In comparison, Telling, Selling, and Participating conditions yielded Phase 2 su ciency in 60 80% of cases, reaching marginal signi cance against baseline yet showing substantial di erences in odds (ORs: 0.03 0.07). (Table 4). Moreover, these three styles uniquely achieved Phase 3 su ciency 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 ndings align with the subjective evaluations (Figure 8). Compared with the baseline, participants perceived signi cantly higher levels of Depth and Su ciency of knowledge co-construction under Telling, Selling, and Participating interventions, while Delegating showed no signi cant di erence.

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 rst 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 o ered 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 owing. 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 di erent agent styles, a contrast we present further in Section 5.2.2.

5.2.2 Di erential Impacts of Intervention Styles. Having examined the e ect of agent presence in general, this section compares how the four styles di erentially 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 Su ciency (Table 4). Despite these statistical similarities, participants perceived the two styles di erently. 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

Table 4: Pairwise comparison results for thread-level measures. Only dyads with at least marginal signi cance (p Ÿ.10) are included. Reported values include the model estimate, standard error (SE), p-values (+: .050 Ÿ p Ÿ.100, *: p Ÿ.050, **: p Ÿ.010, ***: p Ÿ.001) adjusted using the Benjamini Hochberg (BH) correction procedure and e ect 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 nish, 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 signi cant improvement in Phase 2 Su ciency (Table 4) and was perceived as fostering greater Su ciency 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, "I felt so smart that I didn't need to add anything ."

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

Figure 8: Participants' perceptions of knowledge co-construction across conditions. Error bars indicate 95% con dence intervals (*? Ÿ ”050, **? Ÿ ”010, ***? Ÿ ”001). For signi cant pairwise comparisons, test statistics and e ect sizes are reported. For all items, a higher score indicates better performance.

interventions due to minimal presence, as P37 remarked, It didn't a ect our discussion.

5.2.3 Phase-Specific E ects of Intervention Styles. Beyond overall e ects, we examined how each style in uenced speci c 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, re ected, 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 o ering 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 re ection. 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 o ering 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 justi cation of why a point merited further elaboration. Yet the same detailed justi cations 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-e ort 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 instruction. This avoidance might help explain Selling's lower rate of Phase 2 su ciency compared with Telling and Delegating (Table 3). One typical case was S04 (Figure 9): despite ve consecutive Phase-2 interventions by Selling, subsequent replies were either absent or super cial, 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 e ect in this phase, with contributions resembling its overall impact.

Negotiation. In this phase, participants worked toward resolving con icts. 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 di erently, carrying a clear stance on reconciling di ering 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 di erent 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 re ne their positions . when it challenged my stance, it naturally pushed me to respond (P18). This dynamic made Participating particularly e ective 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 reect 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 re ecting 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 re ect, whether through Telling's blunt prompts or Selling's persuasive framing, were largely ine ective in transitioning discussions into the co-construction stage.

5.3 RQ2: How do intervention styles in uence 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 di erences in both subjective and objective measures. Delegating elicited signi cantly fewer human replies compared with the other three conditions (Table 4). Subjectively, it was consistently rated as signi cantly 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 signi cantly 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 o ering 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 di er signi cantly 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 aw. 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 di er signi cantly 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 signi cant di erences across the ve conditions in perceived Effort, Frustration, Complexity, or Agency However, participants reported signi cantly 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 speci c 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 a ect human-human interaction?

While subjective ratings of human human interaction showed no signi cance (Figure 12), thread-level data revealed that Telling elicited signi cantly 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,

Figure 10: Participants' perceptions of agents across styles. Error bars represent 95% con dence intervals (* ? Ÿ ”050 **? Ÿ ”010 ***? Ÿ ”001). For signi cant pairwise comparisons, test statistics and e ect 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 e ect 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 rst 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-speci c insights. For Phase-2 interventions of Telling and Selling, directives or persuasion targeted at speci c human comments lowered the barrier for some to reply to those messages (P3 5, P17, P33). As P17 explained, 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" (P32); 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-o 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 ndings, we rst highlight the e ectiveness of the proposed process-orchestrated design, then discuss its limitations and generalizability. We next examine the tensions between di erent styles and explore potential resolutions, and nally propose key design considerations for future research.

6.1 Process-Orchestrated, Phase-Sensitive Agent Design

The agents adopted a simple but e ective intervention logic: ensure su ciency in the current phase before explicitly pivoting to the next phase. This di ers from prior work that optimizes for single-phase behaviors. For example, the agent proposed by Ito et.al. [69] prompts for elaboration and clari cation within the