

Bridging Talk and Thought: Understanding Dialogue Dynamics Across Collaborative Problem-Solving Contexts

Zhengyuan Liu^{♦*}, Stella Xin Yin^{♦*}, Min-Yen Kan[♦], Nancy F. Chen[♦]

[♦]Nanyang Technological University, Singapore

[♦]Agency for Science, Technology and Research (A*STAR), Singapore

[♦]National University of Singapore

{liu_zhengyuan, nancy_chen}@a-star.edu.sg

Abstract

We present a conceptual framework for analyzing dialogue in collaborative problem-solving contexts, with an emphasis on the emerging dynamics of human-AI and multi-agent collaboration. As intelligent systems become active agents capable of autonomous reasoning and strategic cooperation, understanding the dialogic interaction during collaborative problem solving is increasingly important for optimizing and evaluating such partnerships. Our framework addresses key limitations in current analytical approaches through a hierarchical two-layer coding scheme that integrates cognitive and non-cognitive problem solving with metacognitive regulatory mechanisms. We demonstrate its effectiveness and generalizability across nine datasets spanning multiple domains, and provide insights into how humans and agents coordinate their knowledge, skills, and efforts to solve complex problems, showing in particular that metacognitive regulation can be an essential discriminator of deeper collaboration.

1 Introduction

Collaborative problem-solving (CPS) is a collective process where two or more agents actively engage in sharing, exchanging, and negotiating their knowledge, skills, and efforts to reach shared solutions (Csapó and Molnár, 2017). In CPS, participants work together to achieve goals that extend beyond individual capabilities and require distributed cognition. The central component of effective collaboration is verbal interchange and communication. Through dialogue, collaborators externalize their thinking, negotiate meaning, identify problems, generate and evaluate solutions, and regulate their collective progress (Mercer, 2002). This dynamic interaction serves as both the vehicle and catalyst for successful collaborative efforts, high-

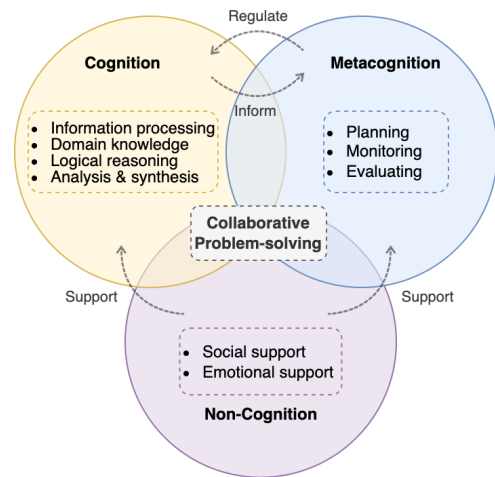


Figure 1: A conceptual model of the dynamic interactions across cognitive, non-cognitive, and metacognitive dimensions during problem-solving.

lighting the essential role of communication analysis in understanding and improving CPS. Humans engage in collaboration to overcome cognitive limitations, leverage diverse expertise, and manage complex tasks that are beyond the capabilities of any individual participant (Dillenbourg, 1999). The distributed nature of collaborative cognition allows groups to process more information, consider multiple viewpoints simultaneously, and generate more innovative solutions than any single person could achieve independently.

Drawing on collaboration theories, successful collaborative communication relies on three interconnected components, as shown in Figure 1: **Metacognition** (regulatory behaviors that monitor and control problem-solving processes) (Stanton et al., 2021), **Cognition** (application of domain knowledge and reasoning) (Mercer, 2002), and **Non-cognition** (socio-emotional expressions that facilitate collaboration) (Bakhtiar et al., 2018; Isohätälä et al., 2020). In this conceptual model, **Metacognition** regulates both individual and col-

* Equal contribution.

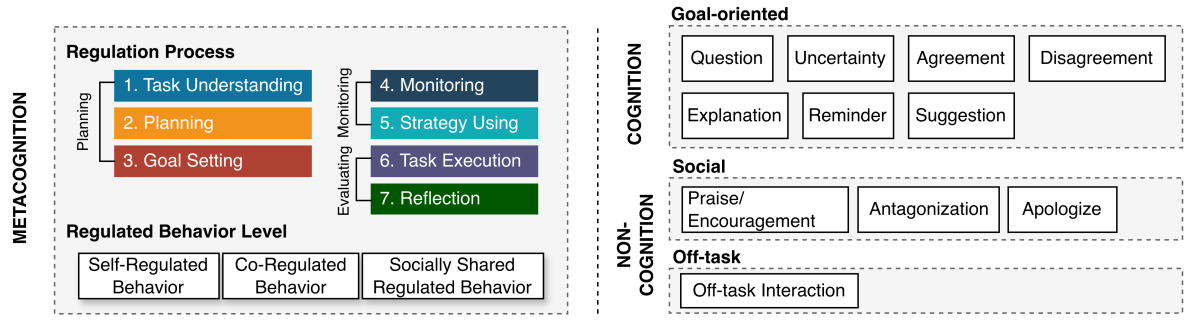


Figure 2: A multi-structure coding scheme for collaborative dialogue analysis. The upper layer focuses on metacognitive processes and regulatory behaviors, capturing both group-level and speaker-level patterns. The lower layer examines the specific cognitive and non-cognitive behaviors at the utterance level.

lective cognitive processes, providing oversight and control functions (Hadwin et al., 2017); **Cognition** provides the information content for metacognitive processes (Dillenbourg, 1999); and **Non-cognition** maintains participant engagement and manages group dynamics in order to create a supportive environment for both metacognitive and cognitive processes to function effectively (Näykki et al., 2014; Bakhtiar et al., 2018).

This conceptual model is particularly critical as human–AI collaboration increasingly extends beyond brief, isolated interactions toward sustained, long-term engagements. Unlike conventional tool-use scenarios lasting only minutes, emerging collaborative tasks may span days or weeks, requiring AI agents to participate in full-pipeline workflows that involve iterative goal negotiation, continuous progress monitoring, and adaptive strategy revision across multiple sessions. In such extended contexts, the dynamic interplay of metacognitive, cognitive, and non-cognitive processes becomes essential not only for task completion but for maintaining shared understanding, sustaining motivation, and preserving joint responsibility over time (Hadwin et al., 2017; Dillenbourg, 1999). However, previous research reveals mixed results regarding the effectiveness of human–AI collaboration. A meta-analysis by Vaccaro et al. (2024) examining over 100 studies found that human–AI combinations, on average, outperformed humans working alone but did not perform better than AI working independently. Research indicates that most human–AI interactions fall into patterns of cooperation or tool use rather than genuine collaboration, such as in medical diagnosis (Reverberi et al., 2022) and data analysis (Jiang et al., 2021), with AI functioning primarily as automated tools that augment human capabilities or as cooperative agents that assist with specific

tasks (Gomez et al., 2025).

Since collaborative dialogue serves as the observable manifestation of collaborative processes, systematic examination of communication patterns could help to identify whether certain interactions contribute to mutual engagement toward shared goals, complementary contributions of unique expertise, adaptive responsiveness to partner actions, negotiated shared understanding, and joint responsibility for outcomes (Galescu et al., 2018; Cohen, 2019; Liu and Chen, 2021; Yin et al., 2025). Conversely, asymmetric dialogue patterns, where one party drives goal-setting, monitoring, and evaluation while the other responds reactively, indicate cooperation or tool use rather than true collaboration. Such understanding can help distinguish and measure the degree of collaboration in human–AI contexts, explain why human–AI collaboration succeeds or fails, and inform the design of systems that achieve genuine collaborative partnerships.

Therefore, in this work, we introduce a multi-structure coding scheme for collaborative dialogue analysis (Figure 2) that characterizes collaborative problem-solving across **Metacognition**, **Cognition**, and **Non-cognition** dimensions. The coding scheme classifies the dialogue patterns at the utterance level and assesses each member’s contribution and distribution during collaboration activities. This approach differentiates collaborative interactions from mere cooperative exchanges or tool use scenarios by identifying when participants demonstrate asymmetry across the three dimensions. We conducted experiments on nine CPS datasets in multiple scenarios. We observed that structured environments often induce functional asymmetry where one participant assumes a metacognitive leadership role, indicating that collaborative patterns are context-dependent and shaped by task

structure. Furthermore, our analysis of human-AI interactions shows that the imbalance remains in common collaborative settings: human users dominate self-regulation, whereas AI systems are inclined to reactive and co-regulated behaviors, which helps explain why human-AI collaboration often fails to achieve genuine partnership and clarifies the current role of AI as a responsive assistant. The framework and coding scheme can be generalized to different domains, particularly, validating metacognitive regulation as an essential discriminator for accurately modeling collaboration. This work also offers practical guidance for developing more synergistic human-AI collaborative systems, such as detecting regulatory imbalance in real time, triggering adaptive scaffolding to balance contributions, and guiding the design of AI agents that can initiate and share metacognitive regulation rather than merely execute assigned tasks.

2 Related Work

Natural language processing research has long focused on modeling dialogue structures and discourse acts to understand communicative intents (Asher et al., 2016; Budzianowski et al., 2018). From statistical to neural models and large language models, various methods are proposed to classify conversational intents (Williams et al., 2013), track dialogue states (Mrkšić et al., 2017; Aksu et al., 2022), analyze discourse (Asher et al., 2016; Liu and Chen, 2021), and evaluate interaction quality in complex, task-oriented contexts (Yin et al., 2025). In linguistics and the learning sciences, collaborative patterns are traditionally analyzed through the lens of exploratory talk and socio-emotional regulation. Foundational studies highlight how shared understanding is dynamically negotiated via epistemic interactions, linguistic alignment, and metacognitive scaffolding (Mercer, 2002; Hadwin et al., 2011; Dillenbourg, 1999). As systems transition from passive tools to active conversational agents, research increasingly evaluates interaction symmetry and joint problem-solving dynamics (Galescu et al., 2018; Gomez et al., 2025; Yin et al., 2025; Tan et al., 2025). Recent works formalize decision-oriented dialogues where humans and AI must combine disparate information to negotiate shared goals (Lin et al., 2024). However, meta-analyses reveal that human-AI pairs often fail to outperform the best individual agent, frequently devolving into asymmetric tool-

use rather than real partnership (Jiang et al., 2021; Reverberi et al., 2022; Vaccaro et al., 2024). To close this gap, recent work has turned to proactive, metacognition-supporting agents (Gmeiner et al., 2025; Mukhopadhyay et al., 2026) and multi-agent collaboration benchmarks (Zhu et al., 2025), seeking to move interaction from cognitive offloading toward shared initiative. In this work, we bridge these computational and cognitive perspectives with a unified coding scheme to capture multi-dimensional interactions underlying collaborative problem-solving.

3 CPS Conceptual Model and Coding Scheme

Collaborative problem solving (CPS) is a collective process where individuals work together to achieve a shared goal. In CPS, two or more agents actively engage in sharing, exchanging, and negotiating their knowledge, skills, and efforts for solutions (Csapó and Molnár, 2017). The communication nature of group-based work highlights cognitive, social, and metacognitive dimensions that dynamically interact during the problem-solving process (Dillenbourg, 1999; Nelson, 2013; Woolley et al., 2010). We introduce a multi-structure coding scheme (Figure 2) to systematically track the above three dimensions: metacognitive awareness through planning, monitoring, and evaluating regulation processes at three behavioral levels; cognitive processes through seven types of goal-oriented interactions; and non-cognitive elements through three types of social interactions. The following subsections explain the theoretical foundations and key components of each dimension in detail.

3.1 Metacognition

Metacognition refers to the process by which participants monitor and regulate their own thinking and behaviors. In collaborative problem-solving contexts, individuals employ knowledge of the task and problem-solving strategies to plan their work, monitor their progress toward goals, and evaluate outcomes (Council et al., 2000; Berliner, 1994). We examine metacognition through two complementary lenses: §3.1.1 regulation processes (planning, monitoring, and evaluating) and §3.1.2 behavioral levels that distinguish how regulation is initiated and maintained (self-regulated, co-regulated, and socially shared-regulated behaviors).

Metacognition	Regulation Process	Definition and Examples Corresponding to the 8-Stage CPS Process
Planning	① Task Understanding	Identifying and defining the problem (Stage I: “What’s going on?”); Understanding what information is already available about the problem (Stage II: “What do we know?”); Investigating root causes of the problem (Stage III: “What are the root causes?”)
	② Planning	Establishing team structure and distributing tasks strategically (Stage IV: “What could we do?”)
	③ Goal Setting	Setting objectives for problem resolution; Determining criteria for selecting the best solution (Stage V: “What’s the best thing to do?”)
Monitoring	④ Monitoring	Tracking progress, summarizing completed work, and identifying remaining tasks
	⑤ Strategy Use	Proposing potential solutions; Evaluating options and selecting the best solutions (Stage VI: “How do we go about it?”)
Evaluating	⑥ Task Execution	Implementing selected solutions (Stage VII: “Have we solved the problem?”)
	⑦ Reflection	Reflecting and making adjustments for continuous improvement (Stage VIII: “Can we improve on what we have done?”)

Table 1: Seven regulation processes in collaborative problem-solving contexts.

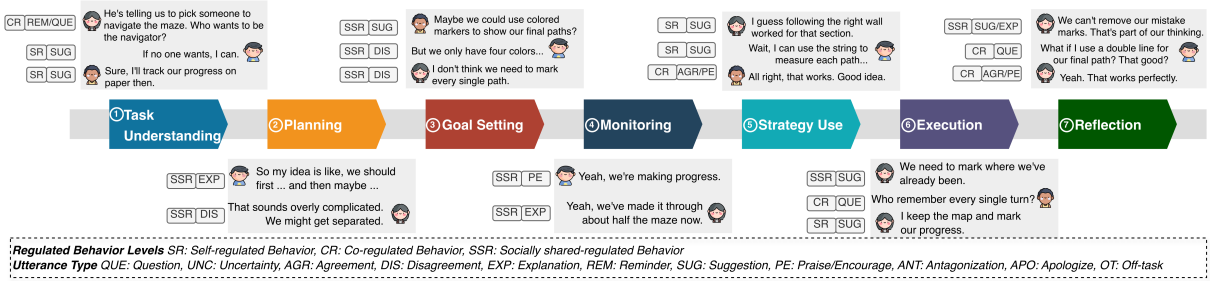


Figure 3: A dialogue example of three individuals collaboratively solving a maze puzzle. The conversation is segmented into seven sequential regulation processes from task understanding to reflection, with each utterance categorized according to regulated behavior levels and cognitive/non-cognitive utterance types.

3.1.1 Regulation Process

Metacognition typically involves planning, monitoring, and evaluating (Stanton et al., 2021). Building on Winne and Hadwin’s (1998) metacognitive model, we decompose it into seven sequential processes (see Table 1). To analyze specific interaction patterns and behaviors within each process, we integrated this framework with the eight-stage CPS process model (Nelson, 2013), developing a coding scheme that examines how groups collectively solve problems through dialogue, from initial problem identification to solution evaluation and refinement. Here, we show a dialogue example (Figure 3) to illustrate how these seven regulation processes unfold in collaborative dialogue. In this example, three students work together to solve a maze puzzle task and identify the shortest route. Their conversation reveals how the team moves through each regulation phase, from initial task comprehension to final reflection on their collaborative process.

Planning ①Task Understanding includes identifying the problem, assessing available information, and investigating root causes. In our maze puzzle

example, this process begins when group members identify problems, as the female student inquires, “*He’s telling us to pick someone to navigate the maze. Who wants to be the navigator?*” ②Planning then builds on this understanding by distributing tasks strategically. One team member suggests, “*So my idea is like, we should first check if there are multiple entrances to the maze, and then maybe split up to try different paths. If we hit a dead end, we can mark it on our map and backtrack...*” ③Goal Setting completes the planning process by establishing specific objectives and criteria for optimal solutions, “*Maybe we could use colored markers to show our final paths? And demonstrate it as the shortest one?*”

Monitoring ④Monitoring refers to tracking progress and identifying remaining tasks. In our example, we observe one member tracks the progress and reminds other members that “*Yeah, we’ve made it through about half the maze now.*” ⑤Strategy Use focuses on proposing solutions from different perspectives, such as: “*I guess following the right wall worked for that section.*”

Evaluating ⑥Task Execution refers to implementing selected solutions, as one participant comments, “*We need to mark where we’ve already been.*” After that, ⑦Reflection involves group members assessing outcomes and reflecting on the process for further improvements. For example, “*We should probably mark our final route more clearly. Like this. So that we don’t confuse it with the dead ends we explored.*”

It is important to note that these seven regulation processes do not operate as a rigid, linear sequence in practice. In complex problem-solving contexts, groups frequently revisit, skip, or cycle through stages in response to emerging challenges and new information. For instance, a team may return from ⑥Task Execution back to ⑤Strategy Use when an implemented solution proves ineffective, or loop back to ①Task Understanding mid-way through ④Monitoring when the group realizes the problem has been misidentified. In our maze puzzle example, this non-linearity is evident when a member’s reflection simultaneously triggers a re-evaluation of their strategy and a revision of their shared goal.

3.1.2 Regulated Behavior Levels

The regulation processes described above focus on *what* problem-solving steps teams follow, while the regulated behavior levels examine *how* group members actually coordinate and work together during collaboration. From speaker perspective, these regulatory behaviors occur at three distinct levels (individual, interpersonal, and collective), depending on who takes the lead in guiding the group’s efforts. Unlike the sequential problem-solving processes that track the team’s progress through tasks, these behavior levels reveal the social dynamics of how participants share responsibility, influence each other, and collectively manage their collaborative work (Hadwin et al., 2017) (see Table 3).

At the individual level, **Self-regulated Behavior** (SR) is often characterized as an active and goal-directed process in which individuals are portrayed as active intentional regulators of their own cognition, metacognition, motivation, and behavior (Pintrich, 2000). In CPS contexts, this manifests when individuals assess their own knowledge gaps, recognize their cognitive limitations, or take personal responsibility for specific task components. Linguistically, self-regulation is marked by first-person singular expressions, with an emphasis on “I” statements (e.g., “*I need to do this,*” “*I don’t understand this concept.*”). In our maze example, “*I’ll track*

our progress on paper then” demonstrates personal responsibility and active metacognitive planning about how individual actions support group goals.

At the interpersonal level, **Co-regulated Behavior** (CR) involves the scaffolding of regulatory processes between group members, where one’s regulatory activity supports or constrains another’s learning and performance (Hadwin et al., 2011). In the CPS context, co-regulation serves critical functions in collaboration because it distributes cognitive load, provides external monitoring when self-regulation fails, and creates opportunities for regulatory skill transfer between participants (Miller and Hadwin, 2015). Beyond the linguistic markers of second-person address (“*Do you think...?*” or “*Can you...?*”), co-regulation involves sophisticated social cognition. For example, in the maze puzzle task, interpersonal regulation happens when one member asks, “*Who remembers every single turn?*” prompting a specific contribution from others, and when another responds to a suggestion with, “*All right, that works. Good idea.*” These instances demonstrate group members are strategically directing cognitive resources and compensating for potential memory limitations in a group.

At the collective level, **Socially shared-regulated Behavior** (SSR) represents the most integrated form of regulation, emerging when group members collaboratively negotiate shared perceptions and goals. This collective regulation involves coordinated strategic planning, joint monitoring of group progress, and collaborative adaptation when needed to optimize performance (Hadwin et al., 2011). Unlike self-regulation and co-regulation, socially shared regulation distributes regulatory responsibility across the entire group, creating a truly collective ownership of the problem-solving process. Linguistically, this form of regulation is characterized by inclusive first-person plural language (“we,” “us,” “our”) and collaborative decision-making expressions. For example, “*What if we set a time limit for each section of the discussion?*” Our maze example includes numerous instances of shared regulation behaviors, such as “*We might get separated.*” and “*I don’t think we need to mark every single path.*” These exchanges show how the group constructs understanding and coordinates its regulatory activities toward shared goals.

We emphasize that all three regulated behavior levels are central to effective CPS. Empirical evidence shows that over-reliance on self-regulation leads to coordination failures, while excessive co-

Interaction	Utterance Type	Definition
Goal-oriented	1. Question (QUE)	Any inquiry or expression of concern seeking clarification, confirmation, or additional information about the task.
	2. Uncertainty (UNC)	Expression of doubt, hesitation, or lack of confidence about ideas, decisions, or information, often indicating a need for guidance or support from others.
	3. Agreement (AGR)	Expression of consensus, acceptance, or alignment with another’s opinion, suggestion, or idea.
	4. Disagreement (DIS)	Expression of dissent, rejection, or misalignment with another’s opinion, suggesting an alternative perspective.
	5. Explanation (EXP)	Providing detailed clarification, elaboration, or reasoning to enhance understanding of concepts or decisions.
	6. Reminder (REM)	Redirecting attention to task requirements, time constraints, or previously agreed-upon goals to maintain focus.
	7. Suggestion (SUG)	Proposing new ideas, approaches, or solutions related to the task.
Social	8. Praise/Encourage (PE)	Offering positive feedback, validation, or motivation to boost team morale and confidence.
	9. Antagonization (ANT)	Making harmful comments, instigating conflict, criticizing personal attributes, or showing irritation toward collaborators.
	10. Apologize (APO)	Acknowledging mistakes, misunderstandings, or inappropriate behaviors with an expression of regret.
Off-task	11. Off-task (OT)	Any utterance unrelated to the current task or collaboration objectives.

Table 2: Cognitive/Non-cognitive utterance types in collaborative problem-solving contexts.

regulation creates dependency relationships that limit group autonomy (Miller and Hadwin, 2015). Conversely, groups that successfully integrate all three levels demonstrate better problem-solving performance and more sustainable collaborative relationships (Isohätälä et al., 2020). We believe that effective CPS requires dynamic interplay among regulatory processes, where each participant must simultaneously maintain their own self-regulation, provide regulatory guidance to others through co-regulation, and participate in collective regulatory efforts through socially shared regulation (Hadwin et al., 2017).

3.2 Cognition

Cognition forms the foundation of the actual thinking processes that team members use to understand problems, generate solutions, and make decisions (Figure 1). This includes analyzing information, applying domain knowledge, and logical reasoning. To understand group members’ cognitive processes from utterance level, we build on Mercer (2002) exploratory talk and previous empirical studies to classify the goal-oriented utterances into seven types (see Table 2 for coding scheme details). They are question (QUE), uncertainty (UNC), agreement (AGR), disagreement (DIS), explanation (EXP), reminder (REM), and suggestion (SUG). These utterance types represent how group members critically and constructively build upon each other’s contributions and create a progressive dialogue that leads to enhanced reasoning and deeper conceptual

understanding (Mercer, 2002). We apply this coding scheme and label each utterance in our maze puzzle example (Figure 3).

3.3 Non-cognition

According to Bakhtiar et al. (2018), social interactions are purposeful interchanges among group members that shape perceptions of emotions and socio-emotional climate within the group. Positive interaction, such as encouragement or positive feedback on group work, has been shown to facilitate productive collaboration (Isohätälä et al., 2020). In contrast, negative interaction, such as discouraging other members from participating or disrespecting other group members, can have adverse consequences on group cohesion, commitment, satisfaction, and performance (Näykki et al., 2014). Although social interactions are not directly related to problem-solving tasks, they are essential requirements for supporting content-related discussion and group cohesion. In CPS, we define three types of utterances: praise and encouragement (PE), antagonistic comments (ANT), and apologies (APO) (see Table 2 for coding scheme details).

In addition to the above dimensions, we also highlight off-task behavior as an indispensable part yet often overlooked component of collaborative dialogue. Off-task interaction refers to utterances unrelated to the current task or objectives, which provide important context for understanding engagement patterns and potential distractions. While off-task interaction may impede problem

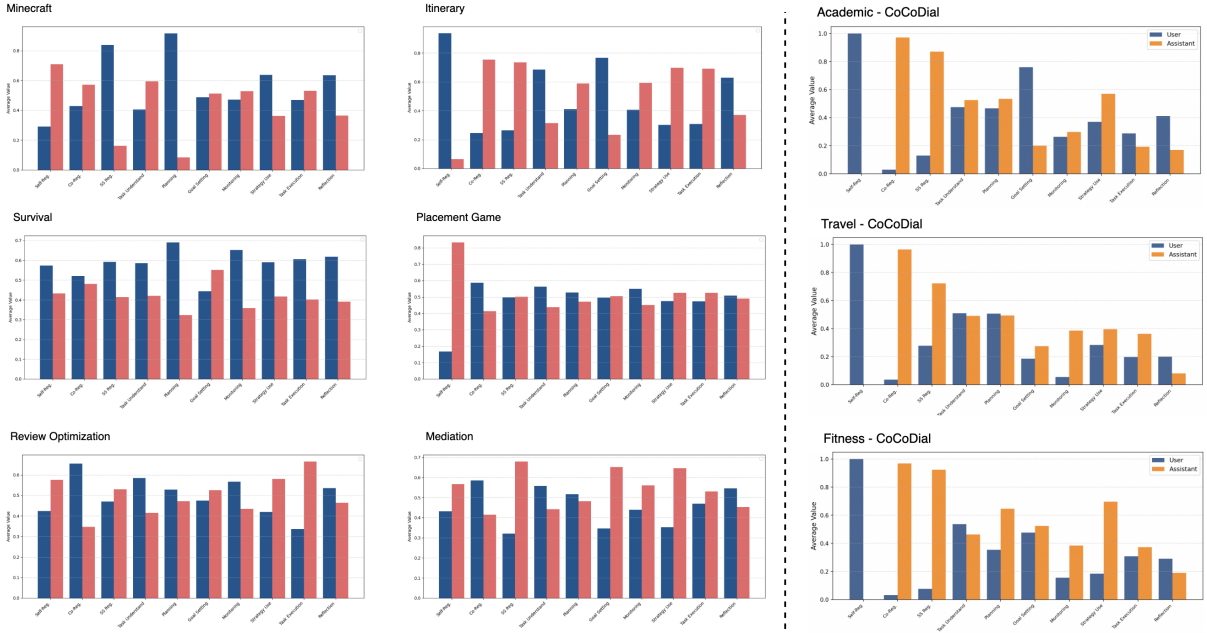


Figure 4: Regulation process characterization in different data: human-human and human-AI. The X axis denotes (from left to right): three regulated behavior levels (*Self-, Co-regulated, and Socially Shared Regulated*), and seven metacognitive regulation processes (*Task Understanding, Planning, Goal Setting, Monitoring, Strategy Use, Task Execution, and Reflection*).

solving, it could have positive roles in CPS if it supports the socio-affective dimensions of interaction (e.g., “*Anyone else surviving on coffee alone today?*” or “*That reminds me of that maze in The Shining—hopefully ours is less terrifying!*”). This argument is supported by previous research (Langer-Osuna et al., 2020) suggesting that there is no clear distinction between ‘off’ and ‘on’ task talk and that group members might weave between ‘on’ and ‘off’ task conversation while remaining engaged with tasks. Such interweaving of interaction contributes to a more productive CPS environment. More importantly, accurately distinguishing between positive and negative off-task behavior cannot be reliably achieved by examining a single utterance in isolation. Therefore, instead of relying solely on the utterance itself, our coding scheme incorporates the surrounding conversational context that considers both the preceding and following utterances to capture the functional role of each off-task turn within the broader dialogue.

4 Collaborative Dialogue Analysis

4.1 Cross-domain Characteristics

Collaborative problem solving manifests distinctively across various contexts, each with unique characteristics that shape interaction patterns and information processing behaviors. As shown in

Table 4, educational collaboration, focusing on knowledge construction and skill development, typically features structured communication within extended timeframes, where process learning is often as valued as outcomes (Roschelle and Teasley, 1995). In workplace settings, CPS is oriented toward organizational goals and problem resolution, characterized by both formal and informal communication channels with strong results orientation (Kauffeld and Lehmann-Willenbrock, 2012). Crisis response contexts involve time-critical, high-stakes situations requiring protocol-driven communication and immediate problem resolution (Kapucu and Garayev, 2011). Game-based collaboration offers another distinctive environment where problem solving is constrained by game rules, often time-limited, with dual focus on process enjoyment and winning (Zagal et al., 2006). Generally, the **explicit roles** refer to formally assigned positions in collaborative tasks (e.g., designated leader, navigator in the maze example), while **embedded roles** emerge naturally through dialogue patterns without formal assignment (e.g., someone consistently monitoring progress or providing encouragement).

4.2 Corpus Selection

We select 6 human-human and 3 human-AI datasets based on our CPS framework: **(D1)** The GAP Corpus (Braley and Murray, 2018) is based on the Win-

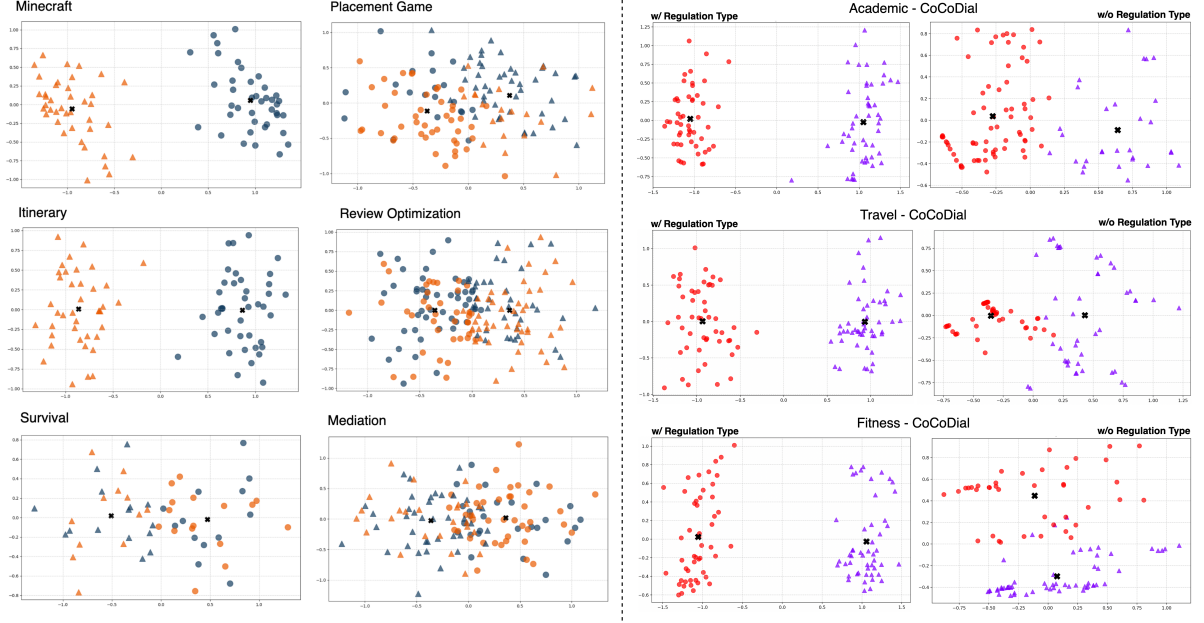


Figure 5: Clustering analysis of collaborative roles in different datasets based on categorization across metacognitive, cognitive, and non-cognitive dimensions. Colors of markers represent explicit roles, while the shapes (e.g., triangle and circle) represent embedded roles within the collaborative dialogue. The clustering patterns reveal how different combinations of communicative elements correspond to distinct collaborative functions.

ter Survival Task, where people rank 15 items by importance in a hypothetical emergent scenario. (D2) Game-based collaboration dataset (Jeknic et al., 2024) contains crowd-sourced interactions between pairs playing a 2D object placement game specifically designed to elicit balanced collaboration. (D3-D5) Three datasets from Lin et al. (2024) include three decision-making tasks of workplace collaboration: paper reviewer assignment, multi-step itinerary planning, and group travel mediation.¹ (D6) MSDC (Narayan-Chen et al., 2019), which provides discourse-annotated collaborative interactions within Minecraft. (D7-D9) Progressive human-AI cognitive alignment in exploratory and personalized scenarios (i.e., academic, travel, and fitness planning).² These datasets were selected based on three criteria: 1) they contain problem-solving interactions with clearly defined goals, 2) they represent diverse contexts, and 3) they include sufficient dialogue data to enable meaningful analysis of collaborative dynamics.

4.3 Dialogue Annotation

For each dataset, we selected 20 representative samples for human annotation. Three trained an-

¹In the mediation scenario, there are three speakers while two speakers play the same role. Therefore, we select the two distinct roles to demonstrate their collaborative characteristics.

²<https://github.com/arenceli/CoCoDial>

notators independently coded each utterance for metacognitive regulation types (Table 1), regulated behavior level (SR, CR, and SSR), and utterance types (Table 2). Inter-rater reliability was calculated using Cohen’s kappa, with disagreement resolved through consensus discussion. Final kappa values ranged from 0.72 to 0.85 across coding dimensions, indicating reasonable agreement.

4.4 LLM-based Utterance Analysis

To efficiently analyze dialogue at scale across multiple datasets, we used GPT-4o for LLM-as-a-judge (Li et al., 2023).³ We created detailed prompts that incorporated our annotation guideline with explicit descriptions and examples of each category (see examples in Table 5). For each utterance in context, the model was instructed to classify: 1) which regulation process the utterance belonged to, 2) which regulated behavior level it represented, and 3) which specific utterance type it exemplified. To ensure reliability, we validated this approach by comparing model classifications against human annotations on a held-out subset of dialogues, achieving substantial agreement with Cohen’s kappa values ranging from 0.69 to 0.75 across the different coding dimensions.

³<https://openai.com/index/gpt-4o-system-card/>

4.5 Regulation Process Characterization

We transformed the utterance-level annotations into speaker-level vectors to capture their collaborative patterns. For each dataset, we first calculate speakers' utterance ratio of each dialogue based on their regulated behavior labels and regulation process labels. We then take the averaged numbers across all dialogues to demonstrate their holistic pattern on the specific CPS task.

Figure 4 presents results on the 9 human-human and human-AI datasets. We found that when participants contribute equally to collaboration, their dialogue exhibits balanced distributions across metacognitive, cognitive, and social dimensions (e.g., Review Optimization and Mediation). In contrast, collaborative interaction in Minecraft and Itinerary shows imbalance in both regulation behavior and processes. This asymmetry extends across multiple metacognitive processes, particularly in Planning, Goal Setting, and Task Execution, suggesting that one collaborator assumes a stronger directive and organizational role while the other adopts a more responsive and executory function. We also observe that in the Placement Game, asymmetry appears when one participant demonstrates significantly higher self-regulated behaviors, while maintaining balanced patterns across other patterns. This sharp contrast suggests that leadership or dominant role in metacognitive regulation can coexist with equitable cognitive contributions. Moreover, on the CoCoDial datasets, we observed that human-AI interactions show a very imbalanced distribution across the three regulated behavior levels, with humans dominating self-regulation while assistants kept Co- and Socially Shared Regulated. This reflects the AI's inherently reactive nature, where responses are always prompted by human input rather than autonomous initiative. When AI seldom initiates planning, goal-setting, or monitoring, the human bears the full regulatory burden (Fan et al., 2025). Detecting this asymmetry is a prerequisite for pushing systems to share regulation rather than passively await instruction (Gmeiner et al., 2025).

These findings have several implications for future collaborative dialogue analysis. The variation in regulatory balance across contexts suggests that effective CPS manifests differently depending on task structure and role configurations. The identification of asymmetric regulation patterns also opens possibilities for real-time collaboration assessment,

where shifts in regulated behavior levels could trigger adaptive scaffolding to rebalance contributions before unproductive collaboration occurs.

4.6 Collaborative Role Clustering

To further study how consistent the speakers' collaborative role is in each CPS task, we performed a clustering analysis on the dialogue-level collaborative pattern. We extract regulation and interaction types from speaker utterances, calculate and vectorize cross-speaker distributions, and then apply K-means to cluster their roles. As shown in Figure 5, the separation of collaborative roles varies significantly based on task structure and the inclusion of metacognitive regulation features. Human-human interactions in the left part demonstrate that highly structured tasks, such as Minecraft and Itinerary, indicate a clear division between speakers. In contrast, tasks requiring more negotiation, like Mediation or the Placement Game, showing overlapped clusters and more symmetrical roles. Moreover, the CoCoDial ablation study on regulation-type labeling demonstrates the critical impact of metacognitive regulation. Including regulation types produces tightly grouped, distinct clusters, while removing them results in highly mixed, diffuse distributions. This confirms that regulation behaviors are essential discriminators for more accurately defining and modeling collaboration.

5 Conclusion

In this paper, we introduced a comprehensive analytical framework that conceptualizes collaborative problem solving as the dynamic interplay of metacognition, cognition, and non-cognition dimensions. Our analysis across diverse contexts reveals that effective collaboration features balanced distributions of metacognitive regulation processes and cognitive contributions. The hierarchical structure of our framework provides a principled approach for evaluating human-human and human-AI partnerships, and points toward a future where both human and AI agents contribute their unique capabilities while maintaining shared agency and responsibility, advancing both theoretical foundations and practical development of effective collaborative systems.

Limitations

All samples used and generated in this work are in English; thus, to apply the model to other

languages, it will require additional data pre-processing steps on the specified language or using multilingual language backbones. We are aware that it remains an open problem to mitigate hallucinations and biases in large language models, which may cause communication issues in human-machine interaction. Of course, current models and laboratory experiments are always limited in this or similar ways. We do not foresee any unethical uses of our proposed methods or their underlying tools, but hope that they will contribute to reducing incorrect system outputs.

Ethics and Impact Statement

We acknowledge that all of the co-authors of this work are aware of the provided ACL Code of Ethics and honor the code of conduct. In our experiments, models are applied under proper license. All data used in this work are only for academic research purposes and should not be used outside of academic research contexts. Our proposed methodology in general does not create a direct societal consequence and are intended to be used to improve the performance, robustness, and safety of AI systems.

Acknowledgements

This research is supported by the Agency for Science, Technology and Research (A*STAR), Singapore, and the National Research Foundation Singapore under the AI Singapore Programme (AISG Award No: AISG3-RPGV-2025-018). We thank the anonymous reviewers for their precious feedback to help improve and extend this piece of work.

References

- Taha Aksu, Zhengyuan Liu, Min-Yen Kan, and Nancy Chen. 2022. N-shot learning for augmenting task-oriented dialogue state tracking. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1659–1671, Dublin, Ireland. Association for Computational Linguistics.
- Nicholas Asher, Julie Hunter, Mathieu Morey, Benamara Farah, and Stergos Afantenos. 2016. Discourse structure and dialogue acts in multiparty dialogue: the STAC corpus. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 2721–2727, Portorož, Slovenia. European Language Resources Association (ELRA).
- Aishah Bakhtiar, Elizabeth A Webster, and Allyson F Hadwin. 2018. Regulation and socio-emotional interactions in a positive and a negative group climate. *Metacognition and learning*, 13:57–90.
- David C Berliner. 1994. Expertise: The wonder of exemplary performances. *Creating powerful thinking in teachers and students*, pages 161–186.
- McKenzie Braley and Gabriel Murray. 2018. The group affect and performance (gap) corpus. In *Proceedings of the ICMI 2018 Workshop on Group Interaction Frontiers in Technology (GIFT)*.
- Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gašić. 2018. MultiWOZ - a large-scale multi-domain Wizard-of-Oz dataset for task-oriented dialogue modelling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5016–5026, Brussels, Belgium. Association for Computational Linguistics.
- Philip Cohen. 2019. [Foundations of collaborative task-oriented dialogue: What’s in a slot?](#) In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, pages 198–209. Association for Computational Linguistics.
- National Research Council, Center for Science, Mathematics, Engineering Education, and Committee on Development of an Addendum to the National Science Education Standards on Scientific Inquiry. 2000. *Inquiry and the national science education standards: A guide for teaching and learning*. National Academies Press.
- Benő Csapó and Gyöngyvér Molnár. 2017. [Potential for assessing dynamic problem-solving at the beginning of higher education studies](#). *Frontiers in Psychology*, 8:2022.
- Pierre Dillenbourg. 1999. What do you mean by collaborative learning? *Collaborative-learning: Cognitive and computational approaches.*, pages 1–19.
- Yizhou Fan, Luzhen Tang, Huixiao Le, Kejie Shen, Shufang Tan, Yueying Zhao, Yuan Shen, Xinyu Li, and Dragan Gašević. 2025. Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2):489–530.
- Lucian Galescu, Choh Man Teng, James Allen, and Ian Perera. 2018. [Cogent: A generic dialogue system shell based on a collaborative problem solving model](#). In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 400–409. Association for Computational Linguistics.
- Frederic Gmeiner, Kaitao Luo, Ye Wang, Kenneth Holstein, and Nikolas Martelaro. 2025. Exploring the potential of metacognitive support agents for human-ai co-creation. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*, pages 1244–1269.

- Catalina Gomez, Sue Min Cho, Shichang Ke, Chien-Ming Huang, and Mathias Unberath. 2025. Human-ai collaboration is not very collaborative yet: a taxonomy of interaction patterns in ai-assisted decision making from a systematic review. *Frontiers in Computer Science*, 6:1521066.
- Allyson Hadwin, Sanna Järvelä, and Mariel Miller. 2017. Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In *Handbook of self-regulation of learning and performance*, pages 83–106. Routledge.
- Allyson Fiona Hadwin, Sanna Järvelä, and Mariel Miller. 2011. Self-regulated, co-regulated, and socially shared regulation of learning. *Handbook of self-regulation of learning and performance*, 30:65–84.
- Jaana Isohätälä, Piia Näykki, and Sanna Järvelä. 2020. Cognitive and socio-emotional interaction in collaborative learning: Exploring fluctuations in students’ participation. *Scandinavian Journal of Educational Research*, 64(6):831–851.
- Isidora Jeknic, David Schlangen, and Alexander Koller. 2024. [A dialogue game for eliciting balanced collaboration](#). In *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 477–489, Kyoto, Japan. Association for Computational Linguistics.
- Jialun Aaron Jiang, Kandrea Wade, Casey Fiesler, and Jed R Brubaker. 2021. Supporting serendipity: Opportunities and challenges for human-ai collaboration in qualitative analysis. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–23.
- Naim Kapucu and Vener Garayev. 2011. Collaborative decision-making in emergency and disaster management. *International Journal of Public Administration*, 34(6):366–375.
- Simone Kauffeld and Nale Lehmann-Willenbrock. 2012. Meetings matter: Effects of team meetings on team and organizational success. *Small group research*, 43(2):130–158.
- Jennifer M Langer-Osuna, Emma Gargroetzi, Jen Munson, and Rosa Chavez. 2020. Exploring the role of off-task activity on students’ collaborative dynamics. *Journal of Educational Psychology*, 112(3):514.
- Minzhi Li, Taiwei Shi, Caleb Ziemis, Min-Yen Kan, Nancy Chen, Zhengyuan Liu, and Diyi Yang. 2023. [CoAnnotating: Uncertainty-guided work allocation between human and large language models for data annotation](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 1487–1505, Singapore. Association for Computational Linguistics.
- Jessy Lin, Nicholas Tomlin, Jacob Andreas, and Jason Eisner. 2024. Decision-oriented dialogue for human-ai collaboration. *Transactions of the Association for Computational Linguistics*, 12:892–911.
- Zhengyuan Liu and Nancy Chen. 2021. [Improving multi-party dialogue discourse parsing via domain integration](#). In *Proceedings of the 2nd Workshop on Computational Approaches to Discourse*, pages 122–127. Association for Computational Linguistics.
- Neil Mercer. 2002. *Words and minds: How we use language to think together*. Routledge.
- Mariel Miller and Allyson Hadwin. 2015. Scripting and awareness tools for regulating collaborative learning: Changing the landscape of support in cscl. *Computers in Human Behavior*, 52:573–588.
- Nikola Mrkšić, Diarmuid Ó Séaghdha, Tsung-Hsien Wen, Blaise Thomson, and Steve Young. 2017. Neural belief tracker: Data-driven dialogue state tracking. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1777–1788, Vancouver, Canada. Association for Computational Linguistics.
- Anirban Mukhopadhyay, Kevin Salubre, Hifza Javed, Shashank Mehrotra, and Kumar Akash. 2026. Exploring the impact of proactive generative ai agent roles in time-sensitive collaborative problem-solving tasks. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, pages 1–25.
- Anjali Narayan-Chen, Prashant Jayannavar, and Julia Hockenmaier. 2019. Collaborative dialogue in minecraft. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5405–5415.
- Piia Näykki, Sanna Järvelä, Paul A Kirschner, and Hanna Järvenoja. 2014. Socio-emotional conflict in collaborative learning—a process-oriented case study in a higher education context. *International Journal of Educational Research*, 68:1–14.
- Laurie Miller Nelson. 2013. Collaborative problem solving. In *Instructional-design theories and models*, pages 241–267. Routledge.
- Paul R Pintrich. 2000. Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of educational psychology*, 92(3):544.
- Carlo Reverberi, Tommaso Rigon, Aldo Solari, Cesare Hassan, Paolo Cherubini, and Andrea Cherubini. 2022. Experimental evidence of effective human-ai collaboration in medical decision-making. *Scientific reports*, 12(1):14952.
- Jeremy Roschelle and Stephanie D Teasley. 1995. The construction of shared knowledge in collaborative problem solving. In *Computer supported collaborative learning*, pages 69–97. Springer.
- Julie Dangremond Stanton, Amanda J Sebesta, and John Dunlosky. 2021. Fostering metacognition to support student learning and performance. *CBE—Life Sciences Education*, 20(2):fe3.

- Bryan Chen Zhengyu Tan, Daniel Wai Kit Chin, Zhengyuan Liu, Nancy F. Chen, and Roy Ka-Wei Lee. 2025. Persuasion dynamics in LLMs: Investigating robustness and adaptability in knowledge and safety with DuET-PD. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 1550–1575, Suzhou, China. Association for Computational Linguistics.
- Michelle Vaccaro, Abdullah Almaatouq, and Thomas Malone. 2024. When combinations of humans and ai are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, 8(12):2293–2303.
- Jason Williams, Antoine Raux, Deepak Ramachandran, and Alan Black. 2013. [The dialog state tracking challenge](#). In *Proceedings of the SIGDIAL 2013 Conference*, pages 404–413, Metz, France. Association for Computational Linguistics.
- P. H. Winne and A. F. Hadwin. 1998. Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, and A. C. Graesser, editors, *Metacognition in educational theory and practice*, pages 277–304. Lawrence Erlbaum Associates Publishers.
- Anita Williams Woolley, Christopher F Chabris, Alex Pentland, Nada Hashmi, and Thomas W Malone. 2010. Evidence for a collective intelligence factor in the performance of human groups. *science*, 330(6004):686–688.
- Stella Xin Yin, Zhengyuan Liu, Dion Hoe-Lian Goh, Choon Lang Quek, and Nancy F Chen. 2025. Scaling up collaborative dialogue analysis: An ai-driven approach to understanding dialogue patterns in computational thinking education. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, pages 47–57.
- José P Zagal, Jochen Rick, and Idris Hsi. 2006. Collaborative games: Lessons learned from board games. *Simulation & gaming*, 37(1):24–40.
- Kunlun Zhu, Hongyi Du, Zhaochen Hong, Xiaocheng Yang, Shuyi Guo, Daisy Zhe Wang, Zhenhailong Wang, Cheng Qian, Robert Tang, Heng Ji, et al. 2025. Multiagentbench: Evaluating the collaboration and competition of llm agents. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8580–8622.

Regulation Types	Self-Regulated Level	Co-Regulated Level	Socially Shared Regulated Level
Task Understanding	"I need to figure out what this problem is asking for."	"Do you agree that we need to understand what's being asked?"	"I think we should also check what resources are already available to us."
Planning	"I'm going to look at the data first."	"Let's split up the work. Maria, could you handle the research part?"	"We could create a shared document to track progress. Does that work?"
Goal Setting	"I will find a solution that reduces costs by at least 15%."	"Our objective should be to reduce costs. Does everyone agree we should aim for a 15% reduction?"	"I think our primary goal should be cost reduction, but we should also consider customer satisfaction."
Monitoring	"I've completed the first part of the analysis."	"Have you finished the analysis section, Sam?"	"I noticed we still need to address the risk section. Should we reassign that or adjust our timeline?"
Strategy Use	"I think approach A is the best option because it addresses our key constraints."	"We could try solution A or B. Which do you all prefer? I think A might work better for our needs."	"A is more cost-effective but B offers better scalability. What if we combine elements from both? Maria, you mentioned a hybrid approach earlier - could we integrate that with these options?"
Task Execution	"I calculated the data based on our plan."	"Can you prepare the slides for reporting our findings?"	"How about we create slides on Google Drive? We might need to adjust our approach based on what we discover during implementation."
Reflection	"I think I can recreate the scatter plot to visualize the result."	"Our solution didn't work as expected. Should we try the alternative method we discussed earlier?"	"I think our solution didn't fully address the problem. Probably because we overlooked the seasonal variations. I remember Sarah mentioned this point before, right?"

Table 3: Examples of different regulation levels in CPS

CPS Context	Definition	Scenario	Role Definition	Outcome Focus
Educational Collaboration	CPS in formal/informal learning environments for knowledge construction & skill development	Knowledge construction focused; scaffolding; emphasis on reflection and metacognition	Teacher assigned or rotated roles	Learning as important as outcome
Workplace Collaboration	CPS in professional settings for organizational goals, innovation, and problem resolution	Efficiency oriented; structured around organizational goals; performance metrics; technology-mediated	Defined by expertise/position	Strong results orientation
Survival Scenario	CPS in time-critical, high-stakes situations requiring rapid coordination	High-stakes decision making; predefined protocols; rapid information processing; minimal deliberation	Highly structured, authority-based	Immediate problem resolution
Game-Based Collaboration	CPS in game environments requiring coordinated problem solving	Engagement driven; artificial constraints; explicit rules; balanced challenges; feedback mechanisms	Player-chosen	Both process enjoyment and winning

Table 4: Collaborative problem-solving contexts and characteristics

Analyze dialogue utterances in collaborative contexts. Classify the utterance into one of three Regulated Behavior types: Self-Regulated Learning (SRL), Co-Regulated Learning (CRL), or Socially Shared Regulated Learning (SSRL).

=====

Regulated Behavior Types:

Type 1. Task Understanding

- Definition: Identifying and defining the problem; Understanding what information is already available; Investigating root causes

- Examples:

* SRL: "I need to figure out what this problem is asking for."

* CRL: "Do you agree that we need to understand what's being asked?"

* SSRL: "I think we should also check what resources are already available to us."

Type 2. Planning

- Definition: Establishing team structure and distributing tasks strategically

- Examples:

* SRL: "I'm going to look at the data first."

* CRL: "Let's split up the work. Maria, could you handle the research part?"

* SSRL: "We could create a shared document to track progress. Does that work?"

Type 3. Goal Setting

- Definition: Setting objectives for problem solving; Determining criteria for selecting the best solution

- Examples:

* SRL: "I will find a solution that reduces costs by at least 15%."

* CRL: "Our objective should be to reduce costs. Does everyone agree we should aim for a 15% reduction?"

* SSRL: "I think our primary goal should be cost reduction, but we should also consider customer satisfaction."

Type 4. Monitoring

- Definition: Tracking progress, summarizing completed work, and identifying remaining tasks

- Examples:

* SRL: "I've completed the first part of the analysis."

* CRL: "Have you finished the analysis section, Sam?"

* SSRL: "I noticed we still need to address the risk section. Should we reassign that or adjust our timeline?"

Type 5. Strategy Use

- Definition: Proposing potential solutions; Evaluating options and selecting the best solutions

- Examples:

* SRL: "I think approach A is the best option because it addresses our key constraints."

* CRL: "We could try solution A or B. Which do you all prefer? I think A might work better for our needs."

* SSRL: "A is more cost-effective but B offers better scalability. What if we combine elements from both? Maria, you mentioned a hybrid approach earlier - could we integrate that with these options?"

Type 6. Task Execution

- Definition: Implementing selected solutions

- Examples:

* SRL: "I calculated the data based on our plan."

* CRL: "Can you prepare the slides for reporting our findings?"

* SSRL: "How about we create slides on Google Drive? We might need to adjust our approach based on what we discover during implementation."

Type 7. Reflection

- Definition: Reflecting and making adjustments for continuous improvement

- Examples:

* SRL: "I think I can recreate the scatter plot to visualize the result."

* CRL: "Our solution didn't work as expected. Should we try the alternative method we discussed earlier?"

* SSRL: "I think our solution didn't fully address the problem. Probably because we overlooked the seasonal variations. I remember Sarah mentioned this point before, right?"

Type 8. None-of-Above

- Definition: The utterance does not present any meta-cognitive behavior.

=====

Give me the type numbers and full type names. Provide your classification with a very concise explanation.

Table 5: One instruction example for collaborative dialogue analysis - Classify the utterance into one of three Regulation types.