SSRLBot: Designing and Developing an LLM-based Agent using Socially Shared Regulated Learning

Xiaoshan Huang*, Jie Gao*, and Haolun Wu

McGill University, Montreal, Quebec, Canada
{xiaoshan.huang, jie.gao3, haolun.wu}@mail.mcgill.ca

Abstract. Large language model (LLM)-based agents are increasingly used to support human experts by streamlining complex tasks and offering actionable insights. However, their application in multi-professional decision-making, particularly in teamwork contexts, remains underexplored. This design-based study addresses that gap by developing LLM functions to enhance collaboration, grounded in the Socially Shared Regulation of Learning (SSRL) framework and applied to medical diagnostic teamwork. SSRL emphasizes metacognitive, cognitive, motivational, and emotional processes in shared learning, focusing on how teams manage these processes to improve decision-making.

This tool paper introduces **SSRLBot**, a prototype chatbot designed to help team members reflect on both their diagnostic performance and key SSRL skills. Its core functions include summarizing dialogues, analyzing SSRL behaviors, evaluating diagnostic outcomes, annotating SSRL markers in conversation, assessing their impact on performance, and identifying interpersonal regulatory dynamics. We compare SSRLBot's capabilities with those of Gemini-1.5, GPT-3.5, and Deepseek-R1 in a case study. SSRLBot demonstrates stronger alignment with SSRL theory, offering detailed evaluations that link behaviors to regulatory dimensions and suggesting improvements for collaboration. By integrating SSRL theory with LLM capabilities, SSRLBot contributes a novel tool for enhancing team-based decision-making and collaborative learning in high-stakes environments, such as medical education.

Keywords: Generative Artificial Intelligence · Large Language Models · Socially shared regulation · Human-Centred Design · Medical Diagnosis.

1 Introduction

Team actions—decisions and operations carried out collaboratively—are essential for achieving success. However, without the joint efforts of team members in adopting appropriate strategies, operations, and regulations, team action alone may not guarantee success [14]. This is especially true in high-stakes decision-making contexts, such as clinical reasoning, where a single decision can lead to

* Equal contributions.

significantly different outcomes [5]. The effectiveness of team decision-making is closely tied to Socially Shared Regulation of Learning (SSRL, [12]), a framework that emphasizes the collaborative regulation of learning processes as key to improving decision-making outcomes. In medical teamwork, team members' interactions through conversation plays a pivotal role in their diagnostic outcome, as they could either direct to the correct or an opposite result [2]. Medical simulation has been pivotal for medical professionals to practice diagnostic skills, and it paves the way for understanding how teamwork works in this field [3,13].

The purpose of this design-based study is to design and develop a large language model (LLM) based agent - SSRLBot, for teamwork evaluation using case conversations in a medical diagnostic task. Grounded from the theory of socially shared regulated learning (SSRL), SSRLBot enables team members to reflect on their diagnostic performance as a team and the key SSRL skills being used to foster team diagnostic success in medical simulation. It also generates instant reports for medical supervisors, providing insights into practitioners' performance and collaboration quality, paving the way for customized scaffolding.

2 Background and Related Work

2.1 SSRL in Team Work

Socially shared regulation is essential in teamwork. It involves team members' mutual efforts in regulating team performance through effective interactions. Socially shared regulated learning reflects learning strategies in educational contexts where two or more learners are involved. Studies in collaborative learning have found that groups that perform better show more SSRL strategies in teamwork [9]. However, the mechanism of SSRL is complex and hard to be captured, as dynamic interpersonal interactions occur where team members are influencing each other. SSRL guides interactions in teamwork from four dimensions, namely meta-cognitive, socio-cognitive, socio-motivational, and socio-emotional aspects. The meta-cognitive aspect of interactions (short in meta-cognitive interaction in the following content) refers to conversations about about the team's cognitive progress [6,11]. Cognitive interaction include conversations about the team's actual cognitive activities, such as planning, performing, and reflecting about the task and team performance [11]. Meta-cognitive interaction to teamwork is like an umbrella to an outdoor walking with weather forecast one, it prevents rains when necessary and can be folded when the sun comes out. It functions as a signal for cognitive actions and monitor important cognitive interactions in teamwork. Socio-emotional interaction, or social emotional interaction is efforts team members made to maintain cohesive and respectful social interaction through team dynamics [8]. Healthy and positive socio-emotional interaction enables team members to engage more effectively in cognitive interaction and lead to better team outcome. Socio-motivational interaction is often discusses with socio-emotional interactions [10,8], as it is pivotal to motivate team spirits and triggers team members to stay tuned and engaged in teamwork.

SSRL research often conducts discourse analysis in computer-supported collaborative learning (CSCL) contexts, as it allows for the identification of learners' situational social, cognitive, emotional, and motivational states through transcriptions of their dialogues or written texts [7]. Integrating trace data from think-aloud protocols and CSCL logfiles represents a significant advancement in capturing learners' thought processes and actions, providing a more objective perspective on SSRL [1]. Most studies analyze verbal data through human coding, which, despite expert evaluation, remains labor-intensive and prone to bias. Large language models (LLMs) learn word usage patterns and apply them to natural language processing tasks [18], offering a more efficient alternative that often matches or surpasses human performance in content understanding, idea generation, and decision-making.

2.2 LLM Agent in Education

LLM agents have recently debuted in educational contexts. For example, PBChat is designed to identify and provide solutions on students' problem behaviors using teacher-parent concersations. LLMAgent-CK [20] is designed as a multi-agent framework to identify middle-school teachers' mathematical content knowledge learning goals from their responses. Other LLM-based agents are designed as virtual tutor or assistants in diverse subjects (e.g., Ruffle & Riley: Biology [15]; Jill Watsons, course readings [17]). Furthermore, other LLM-based agents have been designed for advancing educational research (e.g., Vizchat for visualizing multimodal learning analytics [19]).

Despite previous applications, their potential for medical education remains unexplored. The SSRL theory-driven design could provide valuable insights into teamwork evaluation. Notably, SSRLBot is the first LLM-based agent designed to assess medical teams' SSRL interactions and their impact on diagnostic performance.

3 SSRLBot: System Architecture

Figure 1 shows the entire diagram from data collection to output evaluation, consisting of three key sections: 1) the theory-based development section, 2) the tuning section, and 3) the evaluation section. The architecture of SSRLBot integrates the GPT model (e.g., GPT-4), instructions, and SSRL rubrics. SSRLBot aims to improve the progress of data processing and optimize the quality of the data, reducing human-caused errors in data annotating and diagnosing. In addition, SSRLBot can provide actionable insights for researchers to improve their abilities in assessing and developing SSRL skills in educational, clinical, and collaborative learning environments.

3.1 Design

Since OpenAI launched the GPT store, GPT apps with different purposes have amplified the functionality of ChatGPT, making ChatGPT fully applicable to

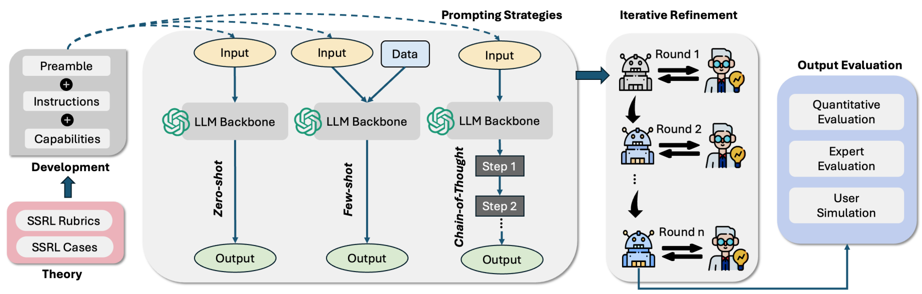

Fig. 1. The Workflow of the System Architecture for SSRLBot.

meet the needs of various users. GPT app configuration consists of four core components, including instructions, knowledge, capabilities, and actions. Instructions refer to development details of the GPT app configuration, such as the app's name, description, logo, instruction description, and conversation starters. Knowledge refers to the external knowledge files that can help improve response accuracy, fairness, and user experience [21]. Users can download these files if the code interpreter is enabled. Capabilities refer to web search, canvas, DALL·E Image Generation, and code interpreter and data analysis. Actions refer to make a third-party APIs to the GPT app. SSRLBot focuses on the instructions and capabilities of the GPT app framework. Furthermore, the SSRL framework is used to support specific functions under the instructions.

3.2 Functional Realization

The instruction description consists of a preamble and function descriptions. The preamble's setting helps SSRLBot clearly understand its role and continue playing this role in a scene for dialogue. We set the SSRLBot as a professional researcher with enough SSRL knowledge and health knowledge, be able to analyze and evaluate data. Moreover, we developed three core functions in SSRLBot: (1) evaluating each team member's interpersonal influence on others, (2) annotating each member's SSRL skills, and (3) providing suggestions for growth.

3.3 Tuning

Although the SSRLBot has been set with three core functions to achieve the goal, the output is not guaranteed to be completely accurate. Tuning plays a key role in building the SSRLBot and enhancing dialogue competencies and accuracy. The tuning section in this study focuses primarily on two parts: instruction and prompt. Instruction tuning helps provide feedback to system development and optimize instructions and preamble. Prompt tuning helps improve input quality. Previous studies [4,16,22] have demonstrated that using more appropriate prompts can improve output quality and mitigate hallucinations. This study conducted four rounds of iterative refinement. The first two rounds focused on

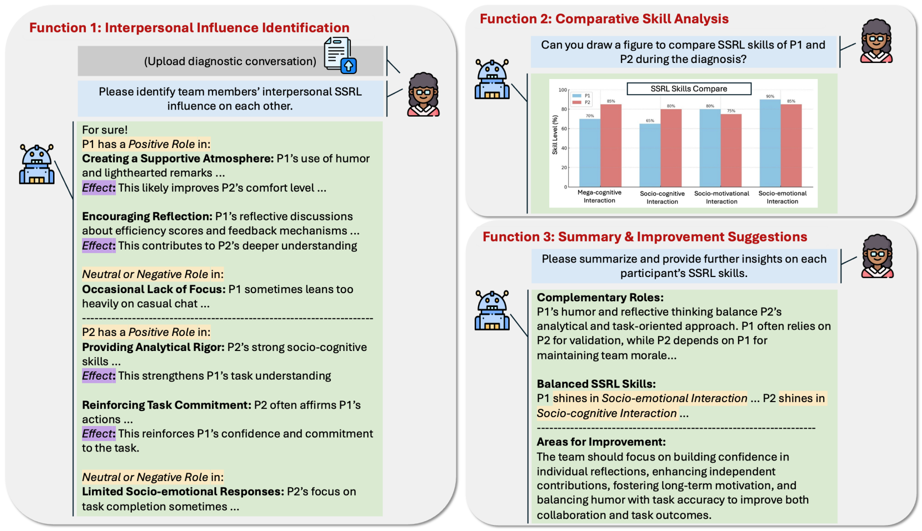

Fig. 2. The Demonstration of SSRLBot Functions.

functional completeness. The third round was mainly adjusted for output accuracy and quality. The fourth round focused on prompt improvement.

4 Illustrative Case Study

4.1 Materials

This section presents a case study using SSRLBot to generate contextualized explanations within the SSRL framework. With ethics approval, we analyzed dialogues from a dyad of medical residents at a large North American university engaged in a team diagnostic task. Figure 2 illustrates functions of SSRLBot.

Specifically, by analyzing the given sample, SSRLBot demonstrates its three core functions. First, it identifies each team member's interpersonal SSRL influence by linking dialogue evidence and summarizing their roles, as shown in Function 1 on the left of the figure. Second, it locates each conversation line and visualizes a comparison of SSRL skills between the two speakers, as seen in the top right corner (Function 2). Third, SSRLBot evaluates the team conversation based on each member's SSRL skills and suggests areas for improvement (Function 3). Together, these functions enable SSRLBot to provide actionable insights into team dynamics, fostering more effective collaborative learning and problem-solving.

4.2 Contextualized Evaluation

For contextualized evaluation, we compare SSRLBot with other LLMs, including Gemini-1.5, ChatGPT-3.5, and Deepseek-R1. The comparisons are guided by: 1.

contextualized description (providing customized details for each human involved in the conversation) and 2. theoretical adherence (alignment with SSRL theory).

All selected LLMs demonstrated their ability to understand the conversation's goal, identify key information, and generate a general report based on the given content. However, ChatGPT 3.5 relied less on contextualized descriptions for each participant, limiting its usefulness for individuals aiming to improve their teamwork skills. Deepseek-R1 provided a more thorough evaluation by blending positive and critical feedback on participants' SSRL skills and their impact on diagnostic performance. However, its theoretical alignment could be improved by explicitly linking behaviors to SSRL dimensions and offering more personalized feedback. Gemini-1.5 provided contextualized descriptions; for example, it acknowledged humor's role within the team and aligned with SSRL theory. However, its analysis lacked actionable feedback for human-centered reflection. In contrast, SSRLBot delivered a more detailed evaluation by categorizing SSRL skills within the metacognitive, cognitive, emotional, and motivational dimensions while highlighting specific conversational examples. Additionally, it provided a comprehensive analysis by offering customized recommendations for each participant, making it clearer where teamwork improvements were needed. By linking evidence from the conversation to each participant's SSRL skills, SSRLBot offered a clearer blueprint for enhancing teamwork, effectively guiding users on which specific dimensions to adapt for improved team decision-making.

5 Conclusion and Future Work

In this paper, we introduced SSRLBot, a knowledge-based LLM agent for teamwork conversation evaluation, grounded from the socially shared regulated learning framework. Integrating the psychological concepts of meta-cognitive, socio-cognitive, socio-emotional, and socio-motivational interactions, SSRLBot aims to facilitate customized feedback for communicators through evaluating their conversations in teamwork contexts. This theory-driven knowledge empowering design enhances LLM functions that support teamwork enhancement through customized feedback for individuals and groups. Our case study showcases SSRLBot's outstanding capabilities: it accesses to every loop of conversation for a comprehensive summary, differentiates each team members' SSRL capacity and their interpersonal influence on each other, visualizes SSRL components with contextually relevant explanations, and offers detailed insights for future growth for teamwork performance. These functionalities can be beneficial for team leaders, educational stakeholders, and team members who want to excel in teamwork tasks through effective communications or interactions.

Future research can enhance SSRLBot by expanding its application beyond dyadic medical diagnosis to diverse contexts, enabling a more comprehensive evaluation of its theory-driven functions. Additionally, while researchers assessed its contextual functions, incorporating participant feedback would provide a more robust measure of its reliability and informativeness.

References

1. Azevedo, R., Harley, J., Trevors, G., Duffy, M., Feyzi-Behnagh, R., Bouchet, F., Landis, R.: Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems. International handbook of metacognition and learning technologies pp. 427–449 (2013)
2. Ball, J.R., Miller, B.T., Balogh, E.P.: Improving diagnosis in health care (2015)
3. Davis, J., Zulkosky, K., Ruth-Sahd, L.A., Frank, E.M., Dommel, L., Minchhoff, D., Uhrich, K.: Health care professional students' perceptions of teamwork and roles after an interprofessional critical care simulation. Dimensions of Critical Care Nursing **40**(3), 174–185 (2021)
4. Diao, S., Wang, P., Lin, Y., Pan, R., Liu, X., Zhang, T.: Active prompting with chain-of-thought for large language models. arXiv preprint arXiv:2302.12246 (2023)
5. Djulbegovic, B., Guyatt, G.H.: Progress in evidence-based medicine: a quarter century on. The lancet **390**(10092), 415–423 (2017)
6. Haataja, E., Dindar, M., Malmberg, J., Järvelä, S.: Individuals in a group: Metacognitive and regulatory predictors of learning achievement in collaborative learning. Learning and Individual Differences **96**, 102146 (2022)
7. Hernández-Sellés, N., Muñoz-Carril, P.C., González-Sanmamed, M.: Computer-supported collaborative learning: An analysis of the relationship between interaction, emotional support and online collaborative tools. Computers & Education **138**, 1–12 (2019)
8. Järvelä, S., Kirschner, P.A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., Koivuniemi, M., Järvenoja, H.: Enhancing socially shared regulation in collaborative learning groups: Designing for cscl regulation tools. Educational Technology Research and Development **63**, 125–142 (2015)
9. Malmberg, J., Järvelä, S., Järvenoja, H., Panadero, E.: Promoting socially shared regulation of learning in cscl: Progress of socially shared regulation among high-and low-performing groups. Computers in Human Behavior **52**, 562–572 (2015)
10. Näykki, P., Järvelä, S., Kirschner, P.A., Järvenoja, H.: Socio-emotional conflict in collaborative learning—a process-oriented case study in a higher education context. International Journal of Educational Research **68**, 1–14 (2014)
11. Nguyen, A., Järvelä, S., Rosé, C., Järvenoja, H., Malmberg, J.: Examining socially shared regulation and shared physiological arousal events with multimodal learning analytics. British Journal of Educational Technology **54**(1), 293–312 (2023)
12. Panadero, E., Järvelä, S.: Socially shared regulation of learning: A review. European Psychologist (2015)
13. Rosen, M.A., Salas, E., Wu, T.S., Silvestri, S., Lazzara, E.H., Lyons, R., Weaver, S.J., King, H.B.: Promoting teamwork: An event-based approach to simulation-based teamwork training for emergency medicine residents. Academic Emergency Medicine **15**(11), 1190–1198 (2008)
14. Salas, E., Sims, D.E., Burke, C.S.: Is there a "big five" in teamwork? Small group research **36**(5), 555–599 (2005)
15. Schmucker, R., Xia, M., Azaria, A., Mitchell, T.: Ruffle&riley: Insights from designing and evaluating a large language model-based conversational tutoring system. In: International Conference on Artificial Intelligence in Education. pp. 75–90. Springer (2024)
16. Song, Y., Wang, T., Cai, P., Mondal, S.K., Sahoo, J.P.: A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. ACM Computing Surveys **55**(13s), 1–40 (2023)

17. Taneja, K., Maiti, P., Kakar, S., Guruprasad, P., Rao, S., Goel, A.K.: Jill watson: A virtual teaching assistant powered by chatgpt. In: International Conference on Artificial Intelligence in Education. pp. 324–337. Springer (2024)
18. Thirunavukarasu, A.J., Ting, D.S.J., Elangovan, K., Gutierrez, L., Tan, T.F., Ting, D.S.W.: Large language models in medicine. Nature medicine **29**(8), 1930–1940 (2023)
19. Yan, L., Zhao, L., Echeverria, V., Jin, Y., Alfredo, R., Li, X., Gaševi'c, D., Martinez-Maldonado, R.: Vizchat: enhancing learning analytics dashboards with contextualised explanations using multimodal generative ai chatbots. In: International Conference on Artificial Intelligence in Education. pp. 180–193. Springer (2024)
20. Yang, K., Chu, Y., Darwin, T., Han, A., Li, H., Wen, H., Copur-Gencturk, Y., Tang, J., Liu, H.: Content knowledge identification with multi-agent large language models (llms). In: International Conference on Artificial Intelligence in Education. pp. 284–292. Springer (2024)
21. Zhang, Z., Zhang, L., Yuan, X., Zhang, A., Xu, M., Qian, F.: A first look at gpt apps: Landscape and vulnerability. arXiv preprint arXiv:2402.15105 (2024)
22. Zhang, Z., Zhang, A., Li, M., Zhao, H., Karypis, G., Smola, A.: Multimodal chain-of-thought reasoning in language models. arXiv preprint arXiv:2302.00923 (2023)