Personalized Learning Artificial Intelligence: Exploring Multimodal Social Robotics in the Classroom

Dr. Aaron Elkins[1,2], Uyiosa Philip Amadasun[1,2,*,†] and Dr. Sabine Matook[3,†]

[1]*San Diego State University, 5250 Campanile Dr, San Diego, California, 92182, USA*

[2]*James Silberrad Brown Center for Artificial Intelligence, 5250 Campanile Dr, San Diego, California, 92182, USA*

[3]*The University of Queensland, School of Business, Colin Clark, 39 Blair Dr, St Lucia QLD 4067, Australia*

Abstract

Social robots, equipped with advanced communication abilities, can now serve as versatile classroom assistants. These robots understand and respond to speech, visual cues, and nonverbal signals, creating dynamic learning environments. Drawing on communication theory, they adapt to various learning styles and work alongside teachers to enhance education. We explore personalized instruction and improved student engagement through multimodal interactions with robot teaching assistants. Our research examines potential benefits, including individualized learning, increased motivation, and support for multilingual and special needs education. We address challenges in visual processing, communication, and cognitive load management. This work advances AI in Education by showing how AI-enhanced robots can create more effective and inclusive learning experiences.

Keywords

Social Robotics, Humanoid Robotics, Education, Artificial Intelligence, Educational Robotics, Communication

1. Introduction

The field of Artificial Intelligence in Education (AIED) has evolved significantly with the emergence of Generative AI and advanced robotics, moving beyond basic service applications to focus on social and human-centric educational settings. The commercial AIED market is projected to reach 4.5 billion pounds in 2024, attracting substantial investments from major tech companies [1].

The integration of Large Language Models (LLMs) with humanoid robots presents promising opportunities for education. Studies demonstrate that advanced Natural Language Processing technologies enhance student engagement and learning [2, 3]. Humanoid robots offer unique advantages over virtual assistants due to their physical embodiment, which leverages the psychological benefits of face-to-face communication [4]. Recent research has extensively explored these educational applications [5, 6, 7, 8, 9].

While LLMs enable more natural, adaptive social interactions in classroom settings, current limitations exist, particularly in processing visual context [10]. This raises critical questions about the minimum capabilities required for effective educational robots. This paper examines how humanoid robot attributes map to goal-directed communication and robot behavior during student interactions, advancing both AIED and potential applications in other social domains like healthcare.

2. Background

Early social interaction research utilized humanoid avatars before physical robot platforms became widely available [11, 12]. Buller and Burgoon [13] characterized communication as a dynamic process

of mutual influence requiring information management, behavioral control, and image maintenance. Effective communication demands managing cognitive load while coordinating verbal and nonverbal behaviors [14], adapting to feedback, and responding to expectation violations [15]. These communication principles must inform humanoid robot design for educational settings. Robots can support teachers by managing classroom interactions and providing individualized tutoring, enabling educators to focus on higher-level instructional objectives while addressing cognitive load and knowledge transfer challenges in the learning environment.

2.1. Classroom Support

Cognitive load occurs when mental resources are consumed by multiple simultaneous tasks [14]. In social interactions, this load can impair an individual's ability to process contextual information and adjust their initial trait attributions based on situational factors. This is particularly relevant as people must process various social cues while regulating their own thoughts and behaviors during interactions.

The concept of cognitive load has important implications for Human-Robot Interaction (HRI) in education. Robots must adapt to diverse learning styles and preferences, which vary dynamically throughout the learning process. The Kolb Experiential Learning Cycle provides a framework for adapting robot communication strategies to student needs [16, 17], consisting of four stages:

1. **Concrete Experience** ("Why?"): Having a specific experience and understanding its relevance
2. **Reflective Observation** ("What?"): Reflecting on observations
3. **Abstract Conceptualization** ("How?"): Theorizing about observations
4. **Active Experimentation** ("What if?"): Applying understanding to new situations

Teachers face significant cognitive load when accommodating multiple learning styles simultaneously, which can impair their ability to adapt teaching methods and maintain effectiveness. Social humanoid robots, enhanced by AI technologies including computer vision, speech recognition, and emotion recognition, can provide cognitive support in the classroom [18]. These robots can assist with personalized learning while teachers focus on classroom management and complex instruction. Through continuous monitoring of student engagement and performance, robots provide actionable insights for teaching adaptation [19], creating an effective feedback loop in robot-student interactions.

2.1.1. Robot-Student Interaction

An important metric the robot must be able to accurately measure and provide feedback to the teacher is the actual success of knowledge transfer from the teacher to the student. Such a metric in itself is multi-faceted. The robot must have a robust and adaptive content delivery system that includes an inference mechanism to detect effective and ineffective knowledge transfer. To develop such a system, we analyzed Buller and Burgoon's schematic model (p.211) of face to face communication to develop what the stages of effective Robot-Student interaction might look like in an educational context.

Figure 1, which was inspired by the schematic model, illustrates a schema of such an interaction. We must state that our derived model is simplified as it cannot encompass all areas contained within this relatively recent phenomenon of humanoid robot-human communication in an educational context. Both the robot and student begin with a pre-interaction mental state shaped by their previous interactions. In this case a previous interaction might encompass: [1] not just a past relations to the current robot, but any other robots which the student might have interacted with, and [2] any priming the student may have undergone to view the robot as an objective authority on knowledge. This will of course affect the students expectations (positive or negative) of the interaction. It has been studied that people find interlocutors of information such as robots or humanoid avatars to be more objective and unbiased, which leads people to open up to them [11]. Such factors affect the students initial receptiveness to receiving knowledge from the robot. The more common factors are of course their preferred learning style and familiarity with the class/course material.
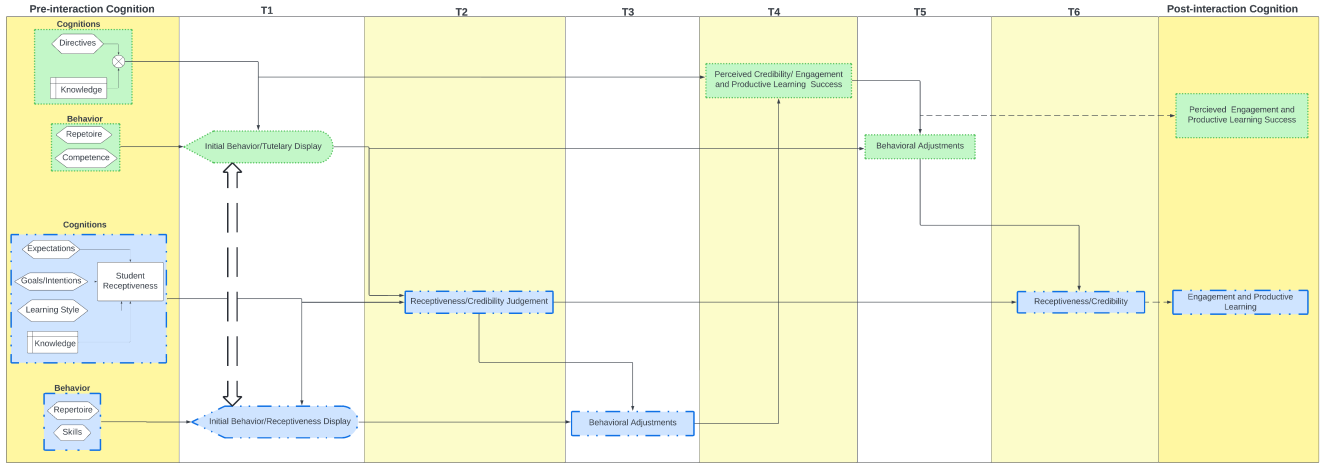
Figure 1: Schema of Robot-Student Communication in a Classroom.

Another factor we take some special interest in are the goals the student may have for the interaction, which we will discuss in our conclusions (see section 4). The robots pre-interaction cognition can be partly categorized as the directives given to the robot - in this case to relay information and provide the teacher accurate measurements of on the effective or ineffectiveness of the relay. Another category can be nominally identified as the robots knowledge. This would include the knowledge stored in any pre-trained and/or fine-tuned models. For example, the knowledge stored in pre-trained LLMs. The knowledge accessible to LLMs via extensions such as Retrieval Augmented Generation or Knowledge graphs (both discussed in section 3), which might include prior interactions with the student, specialized course specific information (such as curriculum), and information about the students learning style.

These pre-interaction cognition states affect the initial behaviors displayed by both parties. The initial receptiveness of the student may manifest itself in body language while the directives and initial knowledge mirror the same mechanism in the robot. It is good to note that these pre-interactional factors may have weight across the entire interaction. The stages between the pre and post-interactional stages can be described as iterative. Just as in Buller and Burgoon's model, these iterative stages can occur simultaneously and in a feedback loop. For simplicity, we describe these stages as a series of credibility checks and behavioral adjustments. The student might voluntarily or involuntarily change displayed behavior based on the perceived credibility of the robots ability to transfer knowledge (or other factors such as comfort). The robot equipped to sense these changes, adjusts behavior accordingly. This might look like an emotionally intelligent robot detecting a change in the students emotional state, which can be inferred from any number of sources (body language, voice modularity, facial expression, etc.) and offering encouragement and motivation. Within these iterative steps the robot continuously measures student engagement. While student engagement has some correlation with effective knowledge transfer, it cannot be the only basis of measurement.

We discern that the robot must be able to draw from different factors - some which are specific to the nature of the learning material - to make some accurate estimate of effective knowledge transfer. The post-interaction stage requires the robot to relay relevant information to the teacher. This will include some permutation of: 1) any major switch in modality of students learning style (as in Kolb's model), and 2) overall engagement and some measurement of how much learning is actually taking place (which would be level of effective knowledge transfer). Now that an overview of the entirety of the feedback process has been defined, let us dive back into the iterative stages.

It is clear that the social robot must have real time adaptive content delivery capabilities which can sense any change in student receptiveness, accurately infer perceived credibility and modulate content delivery method accordingly. As we describe what the robot must be capable of, we are in essence describing some system that provides the robot a sophisticated level of emotional intelligence. Such a system could easily applicable not just in education, but in mental health and other dyadic scenarios.

The robot handles repetitive but dynamic tasks during these iterative stages, including basic question-answering. As a result, teachers can devote their cognitive resources to more complex responsibilities: learner-centered instruction and personalized interventions. In diverse classrooms, these robots can also provide multilingual support, translating or explaining concepts in multiple languages, thus reducing the cognitive load on teachers in diverse educational settings (e.g., multi-language or special needs). The next section discusses our efforts towards the development of an emotionally intelligent system with multi-modal perception, communication features, and affordances.

3. Current Developments and Challenges in AIED with Social Robots

We selected SoftBank's Pepper and NAO humanoid robots [20] for their human-like form and scale, with Pepper's 1.20-meter height providing suitable presence in crowded environments. We developed APIs to enhance their AI and language capabilities for real-time operation, as illustrated below in Figure 2.
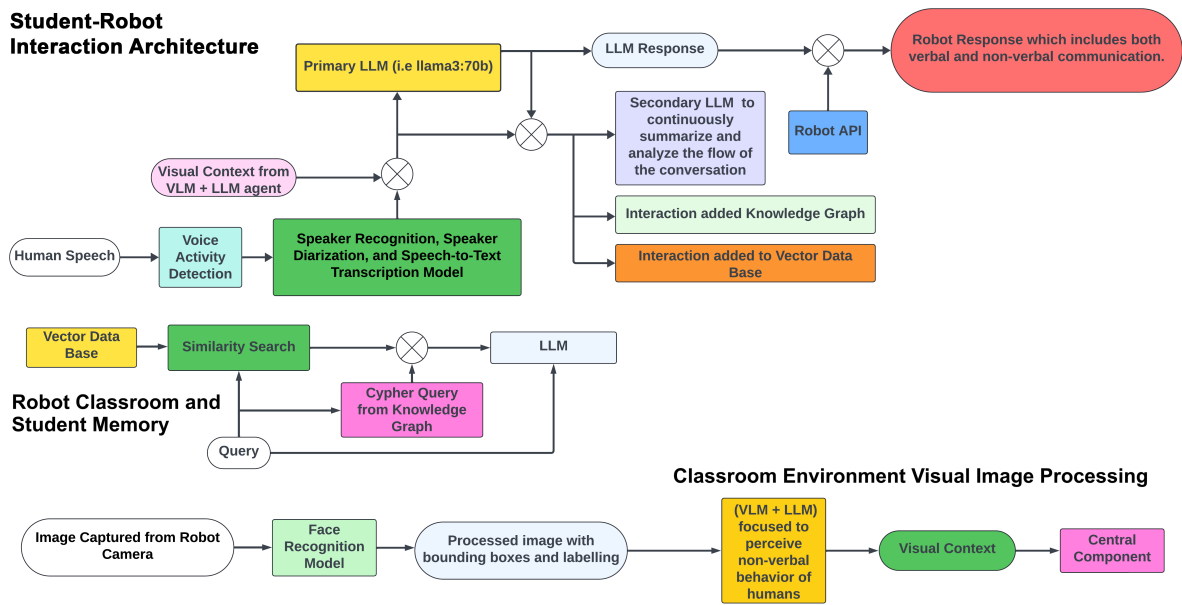
Figure 2: System architecture

The architecture centralizes perceived information to generate robot responses. While utilizing standard processes like voice activity detection (VAD), we enhanced the system with speaker diarization [21] and recognition to distinguish multiple speakers and track long-term student engagement. Robot responses combine spoken LLM output with nonverbal gestures via native APIs. Long-term interactions are stored in vector databases and knowledge graphs, with context-relevant information retrieved through Retrieval Augmented Generation (RAG) [22].

Unlike existing applications that provide static information [23], our system processes dynamic student-robot relationships. Visual information is contextualized using image-to-text Vision Language Models (VLM) to incorporate environmental context into interactions.

3.1. Challenges

Key technical challenges include hardware constraints, model intelligence for real-time processing, and the evolution from reactive to proactive communication [24, 25]. While ongoing work addresses these technical challenges through AI optimization [26], temporal reasoning capabilities [27], and multimodal processing advancements [28], our current implementations demonstrate the viability of AI-enhanced robots in educational settings.

4. Conclusion

This paper explored the potential of integrating AI technologies with humanoid robots to create responsive and adaptive learning environments. Despite technical challenges in real-time processing and proactive communication, our proposed system architecture, combining NLP, computer vision, audio processing, and nonverbal behavior, offers a practical approach to enhancing educational accessibility and support. By addressing challenges such as cognitive load management, visual context processing, and multimodal communication, these AI-enhanced robots can serve as assistants for teachers, adapting to diverse learning styles and individual student needs.

The Robot-Student Interaction model recognizes that each student brings their own unique spectrum of social skills, communication styles, and interpersonal sensitivities to the interaction [29]. Students actively participate in shaping the educational dialogue, contributing valuable information through their distinct communication patterns. While the nature of this interaction varies across educational levels—from primary school to university settings—each student possesses their own sophisticated set of social and communicative competencies. This opens interesting research possibilities for understanding how robot-student interactions might enhance or adapt to different communication styles, and how robotic intervention could support students in developing their natural communication strengths further.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly for grammar and spelling check. The authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] S. Zouhaier, The impact of artificial intelligence on higher education: An empirical study, European Journal of Educational Sciences 10 (2023) 17–33.

[2] T. Wambsganss, A. Janson, J. M. Leimeister, Enhancing argumentative writing with automated feedback and social comparison nudging, Computers & Education 191 (2022) 104644.

[3] R. Winkler, S. Hobert, T. Fischer, A. Salovaara, M. Söllner, J. M. Leimeister, Engaging learners in online video lectures with dynamically scaffolding conversational agents, in: European Conference on Information Systems (ECIS), 2020.

[4] J. M. Lewohl, Exploring student perceptions and use of face-to-face classes, technology-enhanced active learning, and online resources, International Journal of Educational Technology in Higher Education 20 (2023) 48.

[5] G. Kurtz, D. Kohen-Vacs, Humanoid robot as a tutor in a team-based training activity, Interactive Learning Environments 32 (2024) 340–354.

[6] I. Buchem, Scaling-up social learning in small groups with robot supported collaborative learning (rscl): Effects of learners' prior experience in the case study of planning poker with the robot nao, Applied Sciences 13 (2023) 4106.

[7] A. Alam, Social robots in education for long-term human-robot interaction: socially supportive behaviour of robotic tutor for creating robo-tangible learning environment in a guided discovery learning interaction, ECS Transactions 107 (2022) 12389.

[8] I. Buchem, N. Bäcker, Nao robot as scrum master: results from a scenario-based study on building rapport with a humanoid robot in hybrid higher education settings, Training, Education, and Learning Sciences 59 (2022) 65–73.

[9] P. B. Ravari, K. J. Lee, E. Law, D. Kulić, Effects of an adaptive robot encouraging teamwork on students' learning, in: 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), IEEE, 2021, pp. 250–257.

[10] D. Sobrín-Hidalgo, M. Á. González-Santamarta, Á. M. Guerrero-Higueras, F. J. Rodríguez-Lera, V. Matellán-Olivera, Enhancing robot explanation capabilities through vision-language models: a preliminary study by interpreting visual inputs for improved human-robot interaction, arXiv preprint arXiv:2404.09705 (2024).

[11] A. C. Elkins, D. C. Derrick, The sound of trust: voice as a measurement of trust during interactions with embodied conversational agents, Group decision and negotiation 22 (2013) 897–913.

[12] J. F. Nunamaker, D. C. Derrick, A. C. Elkins, J. K. Burgoon, M. W. Patton, Embodied conversational agent-based kiosk for automated interviewing, Journal of Management Information Systems 28 (2011) 17–48.

[13] D. B. Buller, J. K. Burgoon, Interpersonal deception theory, Communication theory 6 (1996) 203–242.

[14] D. T. Gilbert, R. E. Osborne, Thinking backward: Some curable and incurable consequences of cognitive busyness., Journal of personality and social psychology 57 (1989) 940.

[15] J. K. Burgoon, Expectancy violations theory, The international encyclopedia of interpersonal communication (2015) 1–9.

[16] D. A. Kolb, Experiential learning: Experience as the source of learning and development, FT press, 2014.

[17] J. N. Harb, S. O. Durrant, R. E. Terry, Use of the kolb learning cycle and the 4mat system in engineering education, Journal of engineering education 82 (1993) 70–77.

[18] H. Woo, G. K. LeTendre, T. Pham-Shouse, Y. Xiong, The use of social robots in classrooms: A review of field-based studies, Educational Research Review 33 (2021) 100388.

[19] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, F. Tanaka, Social robots for education: A review, Science robotics 3 (2018) eaat5954.

[20] S. Robotics, Pepper the humanoid and programmable robot| softbank robotics, 2021.

[21] D. Garcia-Romero, D. Snyder, G. Sell, D. Povey, A. McCree, Speaker diarization using deep neural network embeddings, in: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2017, pp. 4930–4934.

[22] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, et al., Retrieval-augmented generation for knowledge-intensive nlp tasks, Advances in Neural Information Processing Systems 33 (2020) 9459–9474.

[23] G. Wilcock, K. Jokinen, Conversational ai and knowledge graphs for social robot interaction, in: 2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, 2022, pp. 1090–1094.

[24] Z. Wang, P. Reisert, E. Nichols, R. Gomez, Ain't misbehavin'-using llms to generate expressive robot behavior in conversations with the tabletop robot haru, in: Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction, 2024, pp. 1105–1109.

[25] E. Nichols, D. Szapiro, Y. Vasylkiv, R. Gomez, I can't believe that happened!: exploring expressivity in collaborative storytelling with the tabletop robot haru, in: 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), IEEE, 2022, pp. 59–59.

[26] G. Ioannides, A. Chadha, A. Elkins, Gaussian adaptive attention is all you need: Robust contextual representations across multiple modalities, arXiv preprint arXiv:2401.11143 (2024).

[27] D. Driess, F. Xia, M. S. Sajjadi, C. Lynch, A. Chowdhery, B. Ichter, A. Wahid, J. Tompson, Q. Vuong, T. Yu, et al., Palm-e: An embodied multimodal language model, arXiv preprint arXiv:2303.03378 (2023).

[28] C. Zhang, J. Chen, J. Li, Y. Peng, Z. Mao, Large language models for human-robot interaction: A review, Biomimetic Intelligence and Robotics (2023) 100131.

[29] R. E. Riggio, Assessment of basic social skills., Journal of Personality and Social Psychology 51 (1986) 649.