Human Factors in Robots, Drones and Unmanned Systems, Vol. 179, 2025, 32–42
https://doi.org/10.54941/ahfe1006372

An AI-Based Adaptive Pipeline for Automated Feedback in Immersive Robotics Learning

Mohammadreza Akbari Lor[1], Bhanu Prakash Vodinepally[1], Tisa Islam Erana[2], Bhavleen Kaur Narula[2], Giancarlo Perez[2], Seth Corrigan[3], Shu-Ching Chen[1], Mark Finlayson[2], Biayna Bogosian[4], and Shahin Vassigh[2]

[1]University of Missouri-Kansas City, Kansas City, MO 64110, United States
[2]Florida International University, Miami, FL 33174, USA
[3]University of California, Irvine, CA 92697 USA
[4]Arizona State University, Tempe AZ 85281 USA

ABSTRACT

In this paper, we present a pipeline and framework for the Intelligent Immersive Learning environment for Programming Robotics Operations (IL-PRO), a novel AI-based approach to assess and enhance learner capabilities in an immersive virtual reality (VR) environment. By integrating telemetry data (both continuous and discrete) and speech data, the IL-PRO pipeline evaluates users' motor skills and cognitive understanding to deliver personalized, real-time feedback that links their conceptual understanding with motor skill performance. Telemetry data captures precise physical human-system interactions which are processed and analyzed using Machine Learning (ML) tools to capture and rate motor skill capabilities, while speech data is analyzed using Natural Language Processing (NLP) techniques in concert with a Large Language Model (LLM) to simultaneously assess comprehension and task-related knowledge. These insights are then integrated and used to provide feedback and adapt the learning environment dynamically, tailoring tasks and modules to the learner's specific needs and progress. To demonstrate the feasibility of this approach, we apply the pipeline to a VR task focused on robot acceleration, which emphasizes how motor skills and cognitive understanding work together when learning about inertia in industrial robotic arms. This use case illustrates the pipeline's comprehensive workflow: data collection, multimodal processing of telemetry and speech using machine learning and AI, integration of cognitive and physical insights, and generation of adaptive, real-time feedback. The IL-PRO pipeline framework advances the development of immersive learning systems, enables research on how users combine motor skills with cognition, and enhances skill acquisition in applied training contexts such as robotics.

Keywords: Immersive learning, Artificial intelligence, Machine learning, Large language models, Multimodal analysis, Adaptive feedback systems

INTRODUCTION

The rapid advancement of robotics and automation is transforming architecture, engineering, manufacturing, and other industries, fundamentally reshaping their operations. As a result, the demand for skilled professionals is rising at an unprecedented rate (Bock & Linner, 2016; García de Soto et al., 2022). Traditional robotics training faces several significant challenges that hinder effective skill acquisition. One of the primary obstacles is the high cost and limited availability of physical robotics equipment, making it difficult for many students to gain hands-on experience. Safety risks associated with physical machinery training present another challenge. Industrial machinery and robots, for example, require careful handling due to their moving parts and the potential for unexpected malfunctions. Training on real machinery without prior experience can lead to accidents or damage to expensive equipment. This restricts the extent to which students can experiment freely, thereby limiting the learning process. Other traditional classroom methods often rely heavily on lectures and textbooks, which may not effectively convey the hands-on skills required to operate and troubleshoot physical systems. Additionally, many existing training programs lack interactivity and adaptability, making it difficult to cater to learners with diverse backgrounds and skill levels.

To address these limitations, we introduce the Intelligent Immersive Learning environment for Programming Robotics Operations (IL-PRO), an innovative framework that leverages adaptive learning methodologies, immersive technologies, and game-based learning to enhance education through an adaptive immersive learning pipeline. IL-PRO minimizes the need for physical robotics hardware, reducing financial constraints and safety concerns while improving accessibility. The framework integrates artificial intelligence (AI), including machine learning (ML) and natural language processing (NLP), to automate task delivery and provide real-time, personalized feedback based on learners' progress. This ensures that students receive tailored guidance and reinforcement, optimizing the learning process. A key component of IL-PRO is virtual reality (VR), developed using the Unity game engine, which allows for seamless integration of physics-based interactions, realistic simulations, and dynamic learning environments. Through VR simulations, students can operate and program virtual robots, interact with realistic models of robotic systems, and practice troubleshooting techniques in a controlled yet dynamic setting. This approach enhances hands-on learning without the risks and limitations associated with physical machinery. IL-PRO leverages gamification principles, such as task-based challenges, achievement rewards, and interactive problem-solving scenarios, to enhance engagement and motivation. By integrating these elements, the framework encourages learners to explore, experiment, and master complex robotics concepts. This multi-faceted approach aims to revolutionize robotics education by providing a scalable, safe, and immersive learning experience. Through AI-driven personalization, immersive VR environments, and gamified learning strategies, IL-PRO equips students with the skills and confidence needed to excel in robot operations across various industries.

MOTIVATION AND RELATED WORK

The general field of robotics involves operators and engineers in problems that arise through locomotion. These problems arise whether one is concerned with the movement of a single robotic arm or coordination of movement among multiple robotic systems. Early lessons in planning and controlling movements of robotic systems require students to work toward an integrated understanding and facility with details of motor performances and their own conceptual understanding. Lack of integration of the two can lead to bad decision making and poor performance in practice.

As a result, the design of learning experiences in robotics must reflect this fact: successful operation and control of robotic systems ultimately hinges on the operator's integration of information from multiple sources–the actions they take, their perception of the results of those actions, and their changing understanding. Initially focused on improving performances of pilots, one helpful framework for designing such learning experiences derives from ecological psychology (Gibson, Olum & Rosenblatt, 1955; Gibson, 1977; Gibson, 1979). Within the ecological approach, researchers adopt a tripartite view of learning and performance, emphasizing what are called cycles of action-perception and cognition (Taylor, 2010). Students learn by engaging in cycles of action and perception within a specified environment in order to develop new ways of thinking and new solutions to complex movement problems. Learning by doing in this way means students must have opportunities to learn by doing, and to observe the effects of their interaction with the environment.

This has several implications for learning design in the context of robot control. First, in cases where operators or engineers directly control the movement of a robotic system, it suggests that practice and development of complex motor skills are insufficient on their own. Likewise, conceptual understanding of physics divorced from one's movement solutions is also insufficient. Instead, the ecological approach to teaching and learning robotics control naturally emphasizes integration of conceptual understanding and motor solutions (Taylor, 2010). It also assumes that the two bodies of knowledge can inform one another and aid student learning. Consistent with constructivist and embodied theories of learning, students' movement solutions can inform conceptual understanding (Abrahamson & Sánchez-García, 2016). The converse is also true. Students' conceptual understanding can be leveraged to inform one's movement solutions.

In addition to its use of an ecological framework for teaching and learning, IL-PRO's learning design also incorporates aspects of Dynamic Systems Theory which conceptualizes learners as adaptive systems capable of exploring and mastering complex tasks through iterative discovery. This view is consistent with the action-perception-cognition cycles of behavior that are central to ecological approaches. The dynamic systems perspective shapes the design of learning tasks to encourage exploration, adaptation, and the incremental mastery of skills. By also aligning with contemporary theories of personalized and immersive learning, IL-PRO offers a dynamic educational experience tailored to the unique needs and growth of each learner.

This combined ecological-systems theory approach to learning robot control is novel for the field. The ecological-systems theory approach also guides the design and affordances of the IL-PRO platform. In particular, the ecological-systems theory approach to learning motivates development of immersive VR experiences that are instrumented to track both students' movement solutions as well as their conceptual understanding.

To analyze a participant's motor performance, telemetry data, such as the digital arm's positioning and configurations, can be used in immersive environments to evaluate and enhance motor skills. Research has demonstrated the effectiveness of telemetry in providing real-time feedback during skill acquisition, particularly in medical VR training simulators, where dynamic time series classification has been used for scoring and assessment (Vaughan & Gabrys, 2020). Machine learning algorithms have also been applied to telemetry data to predict skill levels and personalize training interventions, with studies highlighting its role in detecting fine motor skill development and creating adaptive learning environments (Polsley et al., 2022). Additionally, integrating telemetry data with other modalities, such as speech and cognitive assessments, has shown promise in offering a comprehensive evaluation of both motor and cognitive skills. For instance, a fully immersive VR system for skull-base surgery training was developed to combine multimodal data, including motion and force telemetry, to assess surgical skills and provide tailored feedback (Munawar et al., 2024). With the rise of Generative AI and Large Language Models (LLMs), Natural Language Processing (NLP) is transforming education by enabling adaptive learning through personalized feedback, Intelligent Tutoring Systems (ITS), and discourse-driven learner interactions (Grenander et al., 2021). Traditional ITS have faced limitations in adaptability due to their reliance on predefined rules and static responses. Recent research has explored the use of Generative AI to enhance adaptive learning by generating personalized feedback (Grenander et al., 2021; Kulshreshtha et al., 2022; Li et al., 2024; Mejeh et al., 2024). For example, an automatic feedback system was developed to analyze learner discourse, detecting correct and incorrect concepts through relational graphs and neural classifiers (Grenander et al., 2021). Further studies have demonstrated the role of Generative AI in improving adaptive learning environments by enhancing feedback mechanisms and refining ITS capabilities (Kulshreshtha et al., 2022; Li et al., 2024; Mejeh et al., 2024).

IL-PRO PIPELINE AND FRAMEWORK

Our team's approach to designing learning experiences emphasizes the integration of movement solutions and conceptual understanding, shaping key criteria for the IL-PRO platform and learning modules. To be effective, participant experiences must take place in realistic environments that promote learning through both physical movement and evolving conceptual comprehension (see Figure 1). Consequently, feedback and task delivery within IL-PRO need to be personalized based on both participants' movement solutions and their conceptual understanding.

One basis for IL-PRO's personalization of instruction is the system's ability to gather real-time log data. The resulting log files describe key features of students' movement and the resulting movement of the IL-PRO robotic arm. Similarly, IL-PRO is also instrumented to capture, transcribe and automatically evaluate students' verbal responses to questions posed by the IL-PRO system. These two capacities permit the IL-PRO system to leverage traditional ML models in addition to existing LLMs to infer and evaluate students' movement solutions via log-file data, and their conceptual understanding via transcribed speech data.

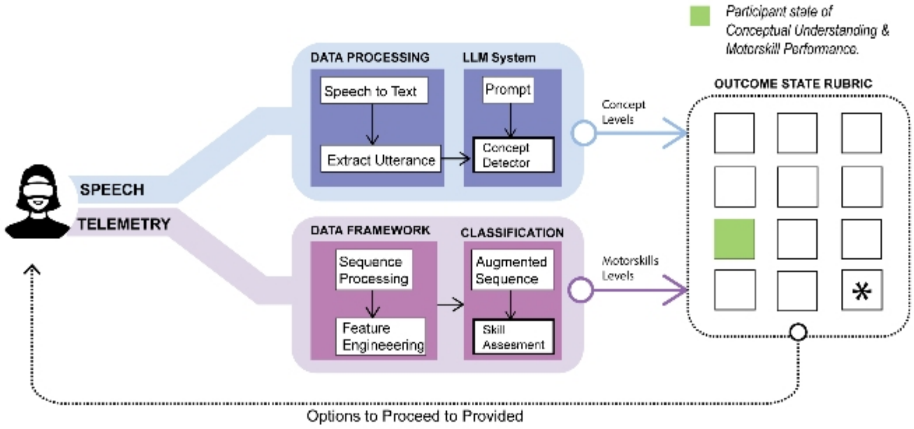

Figure 1: IL-PRO framework architecture.

Telemetry Data

We collect telemetry data, which encompasses both continuous and discrete data types, and incorporate into analysis within our system to analyze user behavior, interactions, and performance. Continuous data consists of real-valued measurements sampled over time, while discrete data represents distinct events or categorical states. Continuous telemetry in VR captures real-time movements and spatial relationships. Depending on the task examples can include: 3D positional data (XYZ coordinates) of objects, rotation angles around the main coordinate axes for object orientations, velocity and acceleration of objects. Discrete telemetry events occur at specific moments, providing context for interactions. Examples include: button presses (e.g., selecting an object, resetting the task), collision events (e.g., hand-object or object-wall interactions), task completion markers, logging timestamps for specific activities, system state transitions, such as scene changes or network disconnections.

We collect telemetry data in VR through the Unity's event-based logging system, capturing key interactions and system changes in real time. Events originate from three primary sources: user actions, environmental changes, and UI interactions. These events are timestamped, assigned unique identifiers, and stored for later analysis by statistical and machine learning methods. The data collection process begins by defining the key variables to be tracked, such as the position and speed of an object, task success or failure, or any other measurable entity. Once identified, a script is attached to the target entity within the VR environment. This script generates a data event based on a timer or state change, depending on the nature of the data. Each data event consists of several parameters: A unique identifier, user ID, session ID, timestamp, event name, data value. These events are transmitted to a central Unity component responsible for collecting all events and assembling a structured file for external processing.

Telemetry data varies widely depending on the task, requiring specialized processing for each application. Continuous data must be handled as a sequence, with different cleaning and feature engineering techniques applied based on the recorded features. For example, when capturing 3D coordinates of an object, movement and time features can be used to derive higher-order derivatives, namely velocity and acceleration. Rotation angles can help calculate deflection angles at each timestamp, while the relative positions of various objects can serve as additional features. We generate these features after data collection rather than integrating them into the data recording process. This approach minimizes computational overhead on the VR hardware, allowing the game engine to record data at a higher frequency. Once extracted, these features provide valuable insights into a participant's motor skills and task performance. Discrete data also plays a crucial role, either as standalone indicators of task state or as auxiliary features that influence how continuous data is processed. Each task requires a custom data analysis and machine learning framework tailored to its specific goals. If a task involves finite states at each step and follows a predefined pattern, a state machine can guide the participant based on their progress. When assessing motor skill performance, movement patterns must be thoroughly analyzed and compared against performance benchmarks. Machine learning techniques, such as skill level classification or expert movement pattern comparison, can help evaluate a participant's proficiency.

Speech Data

To assess a participant's understanding of a given concept, we collect speech data in the form of a conversation between the student and the instructor. The instructor, or pedagogical agent, prompts the participant to engage in dialogue by following a think-aloud protocol, a widely used cognitive strategy where learners verbalize their thought processes while completing a task (Nielsen et al., 2002; Eisenberg et al., 2017). Think-aloud methods are instrumental in educational research, as they provide insights into cognitive reasoning and conceptual understanding in immersive learning environments (Blikstein & Worsley, 2016).

We collect each participant's think-aloud data from recording audio when they engage in a lesson. At the end of a task, they are prompted to answer particular questions or explain reasoning behind their actions with the robotic arm. The questions are set in a way that from the answers, we will be able to detect at what level the student is in understanding the concept. For further processing, we convert the audio of the participant (and of the instructor) to texts using an off-the-shelf automatic speech recognition tool. We segment the think-aloud data into utterances (or sentences) that can be further labeled for concept detection.

We have defined and categorized the participant's levels of conceptual understanding that quantify the amount of completeness of understanding they demonstrate. These levels represent incremental stages in a participant's progress from no understanding of concept at all, to a full understanding. To detect the conceptual level of understanding of the participants from their speech where they think aloud their reasonings, we are following a generative approach using LLMs. To establish a baseline model, we use instructional prompts to categorize the student's levels of understanding given their speech as input text. We use labeled data that categorizes each segmented text according to the participant's levels of understanding, refining the prompt in a few-shot manner by providing examples of participant's utterances along with the corresponding labels. Each example indicates the presence of a concept and, if present, one of four defined levels of understanding: NA (not applicable), L1 (no understanding), L2 (partial understanding), L3 (full understanding), and L4 (application of the conceptual understanding to a specific scenario). To provide the LLM examples specific to our use case for each of the levels of understanding, we have annotated speech of 6 participants for all the tasks and conceptual modules they have participated in. Prior to performing annotations, we developed a codebook outlining the tasks of each module in the curriculum. The codebook contains guidelines that give insights to the annotators about the definitions of each concept and sub-concept in all the tasks in the curriculum, the definition of the four levels of understanding, and specific examples with the student's quoted speech demonstrating the levels of understanding.

Adaptive Feedback Mechanism

The adaptive feedback system integrates telemetry and speech data to provide real-time, personalized feedback. Telemetry data provides quantitative insights into participants' motor-skill performance, such as precision, response time, and control accuracy in the ball-balancing task. Simultaneously, speech data captures qualitative indicators of conceptual understanding, analyzing verbal responses to detect misconceptions, knowledge gaps, or confidence levels. Together, these insights inform the feedback system, allowing it to tailor interventions in real-time.

The feedback framework employs an adaptive decision-making process to prompt participants with targeted feedback after the completion of a task. Depending on their performance and conceptual understanding, participants receive one or more types of feedback: (i) Clarifying Questions, which encourage learners to reflect on their approach and reasoning; (ii) Conceptual Definitions, which reinforce fundamental principles of related to the task where gaps are identified; and (iii) New Challenges or Tasks, which introduce variations of the original task to deepen comprehension and improve skill execution. After receiving feedback, participants can choose to engage with it or skip to the next task, allowing for self-directed learning while maintaining flexibility in progression. By leveraging telemetry and speech-based insights, this framework ensures that feedback is timely, relevant, and reinforcing, optimizing learning outcomes in both motor-skill and conceptual development (see Figure 2).

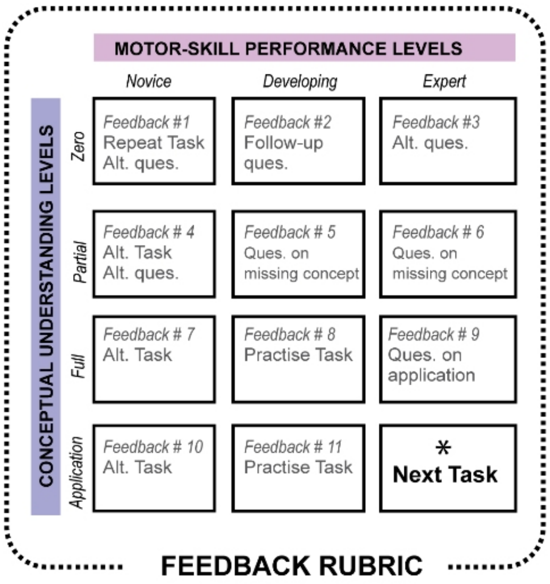

Figure 2: IL-PRO feedback rubric based on student's performance levels.

USE CASE: BALL BALANCING TASK

In this task, participants control a robotic arm within the VR environment. The robotic arm holds a paddle with a ball resting on top (see Figure 3). The objective is to maneuver the robotic arm in such a way that the ball remains balanced on the paddle and does not fall to the ground. The VR environment is designed to realistically simulate the physics of the ball, specifically inertia and gravitational effects. As participants control the robotic arm, they must carefully adjust their movements to maintain the ball's stability, taking into account how gravity along with motion influences its trajectory. The task is progressively challenging, as participants must complete it using three different ball sizes and weights. To successfully complete each attempt, they must guide each ball through a series of predefined waypoints without dropping it. This exercise is designed to help participants develop a deeper understanding of inertia and how it affects motion control. By actively engaging in this task, the expectation is that they can learn the principles of physics related to inertia and practice their motor control strategies, improving their ability to anticipate and adjust for the ball's movement. Through repeat attempts guiding the various balls, students are able to investigate a variety of movement solutions. In particular, through their attempts to complete the task they are able to perceive how the mass of the different balls impact their movement solutions. Repeating the task helps participants refine their movement strategies, making them more adaptable to changes in the ball's mass. This process also enables them to develop a practical understanding of inertia, as they recognize the need to adjust their movements in response to varying mass.

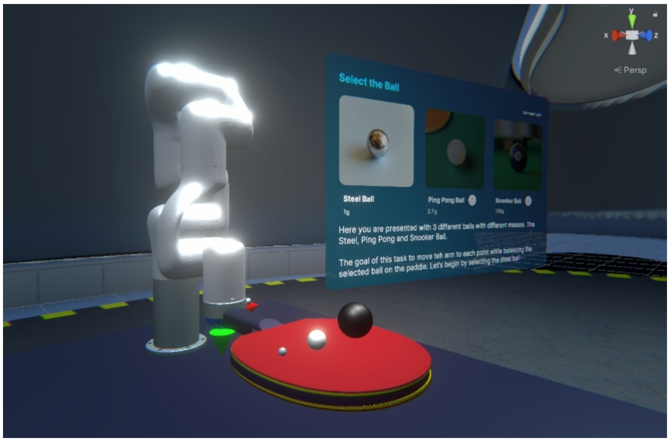

Figure 3: VR screenshot from IL-PRO's ball balancing task use case.

For this task the telemetry data is continuous and is timestamped as recorded. The frequency is 9Hz and the features collected at any instance are the XYZ positions of the sphere, the XYZ position of the paddle and the 3 rotational dimensions of the paddle. Since the task is to go through four waypoints in the play space, the data contains markers for each waypoint reached at any given attempt. whether the attempt fails or succeeds is also marked at the end of each attempt. The data processing framework for the telemetry data is designed so it can systematically prepare and analyze the time-series data derived from VR-based experimental settings involving dynamic interactions between objects such as spheres and paddles. The height of the paddle is used as a reference point to keep track of whether the ball is on the paddle for movement processing and not including the noise data of when the ball has fallen off. As mentioned in the Telemetry Data section, these raw features are processed to compute more features for the proper analysis and modeling. For example, relative spatial metrics such as the distance between the sphere and the paddle are calculated to quantify their interactions dynamically. Since one of the main causes of the sphere movement is the rotations of the paddle along with the movement, the deflection angles for each of the 3 rotation axes are calculated. The cumulative deflection and cumulative movement of the paddle are included to provide a comprehensive view of motion behaviors.

The analytical component of the framework employs Dynamic Time Warping (DTW) to assess the alignment of time-series sequences (Gold and Sharir, 2018). DTW is utilized to compute a cost metric representing the similarity between a participant's motion patterns and predefined reference trajectories recorded by an expert performing the same task multiple times. The more a participant's movement deviates from the expert patterns, the higher the calculated DTW score. This score can be used to directly classify the expertness level of a participant based on statistically derived or heuristic thresholds. Also, by combining all of the above, the resulting dataset which comprises raw and derived features alongside DTW alignment metrics, can serve as a robust foundation for downstream predictive modelling and analysis regarding motor skill assessment from continuous telemetry data. This can be used to train models to perform live motor skill classification for the feedback to the participants.

Throughout the ball balancing task, we want the students to understand the concept of inertia. As discussed in the Speech Data section, the data collection process is the same in this task. Furthermore, we have also hand labeled 2 participant's speech in the particular task of ball balancing where they are taught the concept of inertia with a robotic arm. We collect each student's speech when they are prompted to answer specific questions regarding their experience in completing the task. Based on their answer to the questions, we detect their levels of understanding in the utterance level with our LLM and choose the maximum level they reach to decide how well they understand the concept of inertia.

CLOSING REMARKS

IL-PRO offers a transformative approach to robotics education by integrating adaptive learning, immersive VR, and AI-driven feedback. By combining telemetry-based motor skill assessment with NLP-driven conceptual analysis, the framework is designed to ensure a comprehensive learning experience that enhances both practical and theoretical understanding. Beyond robotics, IL-PRO's principles can be applied to other domains requiring cognitive-motor integration. Future work will focus on refining AI models for deeper personalization and expanding VR-based training scenarios. Through these advancements, IL-PRO aims to improve skill acquisition and better prepare learners for automation-driven industries.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation (NSF) Award No. 2202610: Research on Emerging Technologies for Teaching and Learning Program. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Abrahamson, D., & Sánchez-García, R. (2016). Learning is moving in new ways: The ecological dynamics of mathematics education. *Journal of the Learning Sciences*, *25*(2), 203–239.

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics, 3*(2), 220–238.

Bock, T., & Linner, T. (2016). Site automation: Automated/Robotic On-Site factories. https://mediatum.ub.tum.de/1484319

Eisenberg, J. D., Banisakher, D., Presa, M., Unthank, K., Finlayson, M. A., Price, R., & Chen, S. C. (2017). Toward semantic search for the biogeochemical literature. *In 2017 IEEE International Conference on Information Reuse and Integration (IRI)* (pp. 517–525). IEEE.

García de Soto, B., Agustí-Juan, I., Joss, S., & Hunhevicz, J. (2022). Implications of Construction 4.0 to the workforce and organizational structures. Journal of Construction Engineering and Management, 148(3), 04022002.

Gibson, J. J. (1977). The theory of affordances. *Perceiving, acting and knowing: Towards an ecological psychology/Erlbaum*.

Grenander, M., Belfer, R., Kochmar, E., Serban, I., St-Hilaire, F., & Cheung, J. C. (2021). Deep Discourse Analysis for Generating Personalized Feedback in Intelligent Tutor Systems. AAAI Conference on Artificial Intelligence.

Kulshreshtha, D., Shayan, M., Belfer, R., Reddy, S., Serban, I., & Kochmar, E. (2022). Few-shot Question Generation for Personalized Feedback in Intelligent Tutoring Systems. PAIS@ECAI.

Li, H., Xu T., Zhang, C., Chen, E., Liang, J., Fan, X., Li, H., Tang, J., & Wen, Q. (2024). Bringing Generative AI to Adaptive Learning in Education. https://arxiv.org/abs/2402.14601.

Mejeh, M., & Rehm, M. Taking adaptive learning in educational settings to the next level: Leveraging natural language processing for improved personalization Education Tech Research Dev 72, 1597–1621 (2024).

Munawar, Adnan, Zhaoshuo Li, Nimesh Nagururu, Danielle Trakimas, Peter Kazanzides, Russell H. Taylor, and Francis X. Creighton. "Fully immersive virtual reality for skull-base surgery: surgical training and beyond." International journal of computer assisted radiology and surgery 19, no. 1 (2024): 51–59.

Nielsen, J., Clemmensen, T., & Yssing, C. (2002). Getting access to what goes on in people's heads? Reflections on the think-aloud technique. *Nordic Conference on Human-Computer Interaction*, 101–110.

Omer Gold and Micha Sharir. 2018. Dynamic Time Warping and Geometric Edit Distance: Breaking the Quadratic Barrier. ACM Trans. Algorithms 14, 4, Article 50 (October 2018), 17 pages. https://doi.org/10.1145/3230734

Polsley, S., Powell, L., Kim, HH. et al. Detecting Children's Fine Motor Skill Development using Machine Learning. Int J Artif Intell Educ 32, 991–1024 (2022). https://doi.org/10.1007/s40593-021-00279-7

Taylor, J. G. (2010). The Perception-Conceptualisation-Knowledge Representation-Reasoning Representation-Action Cycle: The View from the Brain. In *Perception-Action Cycle: Models, Architectures, and Hardware* (pp. 243–285). New York, NY: Springer New York.

Vaughan, Neil, and Bogdan Gabrys. "Scoring and assessment in medical VR training simulators with dynamic time series classification." Engineering Applications of Artificial Intelligence 94 (2020): 103760.