education sciences

MDPI

Review

Generative AI-Based Platform for Deliberate Teaching Practice: A Review and a Suggested Framework

Yehudit Aperstein [1], Yuval Cohen [1,*] and Alexander Apartsin [2]

[1] Department of Intelligent Systems, Afeka Tel-Aviv College of Engineering, Tel-Aviv 69988, Israel; apersteiny@afeka.ac.il
[2] School of Computer Science, Faculty of Sciences, HIT-Holon Institute of Technology, Holon 58102, Israel; alexanderap@hit.ac.il
* Correspondence: yuvalc@afeka.ac.il; Tel.: +972-545398152

Abstract: This paper begins with a comprehensive review of the deliberate teaching practice literature related to generative AI training platforms. It then introduces a conceptual framework for a generative AI-powered system designed to simulate dynamic classroom environments, allowing teachers to engage in repeated, goal-oriented practice sessions. Leveraging recent advances in large language models (LLMs) and multiagent systems, the platform features virtual student agents configured to demonstrate varied learning styles, prior knowledge, and behavioral traits. In parallel, mentor agents—built upon the same generative AI technology—continuously provide feedback, enabling teachers to adapt their strategies in real time. By offering an accessible, controlled space for skill development, this framework addresses the challenge of scaling and personalizing teacher training. Grounded in pedagogical theory and supported by emerging AI capabilities, the proposed platform enables educators to refine teaching methods and adapt to diverse classroom contexts through iterative practice. A detailed outline of the system's main components, including agent configuration, interaction workflows, and a deliberate practice feedback loop, sets the stage for more personalized, high-quality teacher training experiences, and contributes to the evolving field of AI-mediated learning environments.

Keywords: teacher training; deliberate practice; generative training; virtual students; large language models; LLM; simulation; personalized instruction

check for updates

Academic Editor: Yu-Chun Kuo

Received: 2 February 2025
Revised: 26 February 2025
Accepted: 19 March 2025
Published: 24 March 2025

Citation: Aperstein, Y., Cohen, Y., & Apartsin, A. (2025). Generative AI-Based Platform for Deliberate Teaching Practice: A Review and a Suggested Framework. *Education Sciences*, *15*(4), 405. https://doi.org/10.3390/educsci15040405

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Training effective teachers is a complex and time-consuming task. The famous "10,000 h rule" could be used to reference teacher training duration. The "10,000 h rule" states that the key to achieving expertise in any domain is a sufficient amount of practicing with at least 10,000 h as a common rule of thumb for time spent on the activity to reach the mastery level in it (Gladwell, 2008; Nichols et al., 2018; Cranford, 2022). However, not all practices are equally helpful in developing a desired set of skills (various teaching and pedagogical skills in this context). Deliberate practice (Ericsson et al., 1993; R. H. Chen, 2022) is a high-quality practice that includes goal-directed activities while focusing on improving one's weaknesses with appropriate feedback from a high-quality mentor.

Unfortunately, acquiring sufficient experience for a teaching professional is very slow. Using some estimates, 5000 h of teaching practice amounts to 5–6 years of teaching experience (Cuban, 2010; Stich & Cipollone, 2021). Moreover, these experiences are rarely deliberate as most of them are not supported by feedback from qualified mentors and do not include repetitive exercises for teaching specific material to students or a class

with certain backgrounds or learning profiles. In addition, unlike many other professions, gaining experience through trial and error might have a significant cost for students in the class of a less experienced teacher. Therefore, there is a need for a drastic increase in the quantity and quality of teaching practice during teachers' training and the initial induction period (Ingersoll & Strong, 2011). There is also a need to provide teachers with new digital applications that will be added to the traditional tools they apply in their teaching (Abedi, 2023).

This paper describes a new conceptual framework for a practical teaching simulation platform using recent advances in simulating group dynamics (Y. Li et al., 2023) and expert feedback (Sonnemann & Blane, 2024). These advances were enabled by the advent of large language models (LLMs), generative AI (Soliman et al., 2024), and collaborative agents. In this platform, students will be represented by a collection of large language model (LLM)-driven agents primed with specific behavioral traits (Jiang et al., 2023), learning patterns, and prior knowledge. The platform will also contain mentor LLM agents to monitor class interaction and provide feedback on teacher performance and helpful improvement suggestions.

Using this platform, aspiring and young teachers could practice deliberately while polishing their skills through the repetitive teaching of specific subjects to students and classes of controlled compositions, behavioral traits, and backgrounds. The mentor LLM agents will monitor the class dynamics, difficulties, and questions asked by virtual students and suggest possible improvements in teaching approaches (Marcus et al., 2020).

During the previous decade, virtual reality was considered the cutting edge of technology, and its use had spread considerably in training and teaching (Malik et al., 2024). However, its development costs were far beyond the reach of a large part of the education system (Extremera et al., 2020). Thus, a more affordable solution was required (Urueta, 2023). The recent advent of LLMs, their capabilities, availability, and price make them an attractive solution for classroom simulations (Lyu et al., 2024). In particular, the ability of LLMs to play roles and personalities enables them to supply, in the form of agents, an environment of chosen personas, including students and mentors (Alsafari et al., 2024). These abilities were recently used to simulate a class environment (Y. Zhang et al., 2024; B. Hu et al., 2024) and to leverage LLM-driven multiagent systems to construct an AI-augmented classroom (Yu et al., 2024). Finally, B. Hu et al. (2024) recommended the future development of a teacher training platform.

Thus, according to the recommendation of B. Hu et al. (2024), we propose a teacher training platform that uses an LLM-based multiagent environment that simulates various students in a class environment and mentors outside the class. The teacher's encounter with these personas is integral to gaining teaching experience and improving teaching capabilities.

This paper continues as follows: Section 2 reviews the literature; Section 3 describes the proposed framework; Section 4 describes a use case; Section 5 analyses the use case and the framework and draws some conclusions and practical takeaways; and Section 6 concludes the paper.

2. Literature Review

To set the stage for our framework, we examine the emergence of LLMs and multi-LLM agents, discuss simulation-based practice, explore tools for modeling and constructing the simulation platform, and address how simulation teaching models can be evaluated.

2.1. The Advent of LLMs and Their Capabilities

There has been impressive progress in generative AI recently, specifically in models for natural language understanding and dialogue. Modern large language models (e.g., ChatGPT 4) surpass the grades of an average student in many standardized tests (OpenAI, 2023), including SAT math, reading and writing, and even the Uniform Bar Exam.

Table 1 presents GPT-4's final score graded on standardized exams, as well as the percentile of test-takers achieving GPT-4's score compares GPT-4's performance to human students on standardized exams, highlighting how generative AI now matches or surpasses human abilities in several fields.

Table 1. GPT performance on standardized exams (based on data from OpenAI, 2023).

Exam	GPT-4	GPT-4 (No Vision)	GPT-3.5
Uniform Bar Exam (MBE+MEE+MPT)	298/400 (~90th)	298/400 (~90th)	213/400 (~10th)
LSAT	163 (~88th)	161 (~83rd)	149 (~40th)
SAT Evidence-Based Reading & Writing	710/800 (~93rd)	710/800 (~93rd)	670/800 (~87th)
SAT Math	700/800 (~89th)	690/800 (~89th)	590/800 (~70th)
Graduate Record Examination (GRE) Verbal	169/170 (~99th)	165/170 (~96th)	154/170 (~63rd)

Moreover, the rate of improvement in reasoning and text generation abilities is impressive, with additional modalities such as voice, images, and videos being integrated and refined. Although state-of-the-art models excel in English, models focusing on other languages are quickly catching up. Significant effort was invested in designing an efficient way to customize generic LLM models for specific domains via fine-tuning (Han et al., 2024), adapters (Z. Hu et al., 2023), or Retrieval-Augmented Generation (RAG) methods (J. Li et al., 2024).

In addition to standardized exams, LLMs have made significant strides in tasks like the Turing Test, which evaluates a machine's ability to exhibit intelligent behavior indistinguishable from a human (Sejnowski, 2023). While no model has fully "passed" the Turing Test in its original formulation, modern LLMs have come remarkably close in many conversational contexts. For instance, GPT-4 and similar models can engage in nuanced, context-aware dialogues that often deceive human judges into believing they are interacting with another person (Bubeck et al., 2023). This capability is particularly evident in customer service, therapy chatbots, and educational tutoring, where LLMs can provide responses that are not only coherent but also empathetic and contextually appropriate (Z. Zhang et al., 2024). These advancements suggest that LLMs are approaching a level of conversational sophistication that challenges traditional notions of human uniqueness in language and reasoning.

In the context of this research, fine-tuning models on a text corpus describing pedagogical concepts might help to improve mentor agents for providing more accurate feedback on teaching approach and class dynamics. Another line of research explores how to make models unlearn selected facts and concepts (Eldan & Russinovich, 2023; Si et al., 2023). While this work provides a different motivation (i.e., copyright concern), the same techniques can be used to produce student LLM models that lack specific knowledge. This will allow an objective evaluation of the teaching strategy based on test performance before

and after the teaching session. Additionally, LLM models that lack specific knowledge can provide valuable insights into areas where teaching strategies may need improvement.

2.2. Multiple Collaborative LLM Agents

Multiple LLM agents collaborate in natural language in multiagent frameworks to achieve a common goal. Usually, each agent is primed with a specific role, background story, available documents, and skills. Naturally, this development has fueled the line of research dealing with simulating human behavior and dynamics by a committee of LLM-based agents. Inspired by the rapid progress of LLMs, an even more powerful multiagent approach has emerged (Park et al., 2023). In the groundbreaking work by Park et al. (2023), a community of 25 agents was released in a virtual town inspired by The Sims game. Each agent's identity, including occupation and relationships, was described in a one-paragraph natural language seed memory. In a short period of time, the agents demonstrated believable social interactions and behavior. Since this research, simulating human behavior with LLMs has become a mainstream research direction, producing a steady and increasing flow of research papers, experiments, and software libraries for developing LLM-based multiagent applications for educational purposes (Ren et al., 2024; La Cava et al., 2024; Zhou et al., 2023; Zhao et al., 2023; Törnberg et al., 2023).

2.3. Simulation-Based Practice

Simulation-based practice is a well-established educational strategy for developing professional competency in a wide array of fields—such as teaching, healthcare, aviation, and military training—by offering realistic, controlled environments that foster experiential learning and facilitate iterative skill refinement (Badiee & Kaufman, 2015; Mikeska et al., 2021; Ledger et al., 2025; Hou et al., 2024). Early attempts at developing education-based simulation leveraged virtual reality technologies with manually defined behavioral scripts for digital simulation (Delamarre et al., 2021; Landon-Hays et al., 2020; Ledger et al., 2025). These systems still use well-trained human actors playing various roles. The advent of avatars and LLM agents generated a new generation of replacements for these well-trained human actors. Numerous efforts exist to create software-based simulation platforms for professional learning (Agyei, 2024; Zheng et al., 2024; Gaviria et al., 2024). Agyei (2024) reports the experience and perspectives of prospective teachers. Zheng et al. (2024) describe a simulation teaching platform for intelligent manufacturing, Gaviria et al. (2024) report the use of simulation for enhancing accounting teaching in higher education, and Sonnemann and Blane (2024) compare three established leading simulation systems. Table 2 compares three digital simulation platforms for teacher training: SimTeach, Proxima, and Teacher Moments.

Table 2. Comparison of three digital simulation platforms for teacher training *.

	SimTeach	Proxima	Teacher Moments
Type	Semi-immersive	Low-immersive	Low-immersive
Established and used	USA, established 2012 Currently used in Australia	UK, established 2022 Currently used in the UK	USA, established 2018 Currently used in the USA mostly
How it works	Virtual student avatars controlled by human actors (human-in-the-loop); interactive conversations; can simulate students, parents, and colleagues.	Text-based scenarios with multiple choice, free text, and voice recording response options; the trainee teacher is given a scenario and then responds.	Text, image, and video-based scenarios with multiple choice, free text, and voice recording response options; the trainee teacher is given a scenario and then responds.

* Data summarized from Sonnemann and Blane (2024).

SimTeach is semi-immersive, using virtual student avatars controlled by human actors for interactive conversations. At the same time, Proxima and Teacher Moments are low-immersive, relying on text, voice, and video-based scenarios with multiple response options. Each platform offers different levels of immersion and interactivity to help trainee teachers practice their skills in simulated classroom environments. However, these approaches are limited to a relatively small number of predefined interaction scenarios and do not allow for personalization to create an intensive, deliberate practice experience as described above. Table 2 summarizes their comparison. In fact, all simulations prior to the recent emergence of commercially available LLMs in 2022, suffer from these drawbacks. However, they were still found to be valuable for the enrichment of teachers' professional training as well as benefitting their professional development (Levin & Muchnik-Rozanov, 2022).

However, the advent of LLMs and their usage in the past two years has enabled a drastically improved simulation. The rapid advancement of generative artificial intelligence enables the creation of more flexible and natural simulation environments for personalized teacher education based on deliberate practice and immediate feedback from AI-based mentors.

Very few papers deal with simulation that uses LLM-based agents and even less so for teacher training (Bhowmik et al., 2024). Bhowmik et al. (2024) present an LLM-powered virtual student conversation agent that they developed for pre-service teacher training in a virtual environment. They found that the simulation with these agents enabled personalized, adaptive training and promoted a more engaging and immersive learning experience for pre-service teachers.

Lee et al. (2023) report about their LLM-based agents: "this research indicates that integrating generative agents into teacher training simulation can be an effective way to offer pre-service teachers with more practical experiences to apply theories and concepts to simulated teaching practices".

2.4. Methodologies for Modeling and Building the Simulation Platform

In software development, "Design Thinking" is a standard methodology for generating personas that simulate different user types. While this approach helps create generalized user models, developing student personas in educational simulations requires more scientifically grounded frameworks. In this context, well-established pedagogical concepts and tools provide the basis for constructing realistic and nuanced student profiles. These studies provide the foundation for building profiles, allowing the simulation platform to accurately capture the complexity of real-world learning environments.

Taxonomies of Learning-Related Cognitive Processes provide a foundational understanding of how student cognition can be modeled. Bloom's Revised Taxonomy (Forehand, 2010; Chandio et al., 2021) categorizes learning into six major cognitive processes: Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating. These processes operate across four knowledge dimensions: Factual Knowledge, Conceptual Knowledge, Procedural Knowledge, and Metacognitive Knowledge. The application of these categories depends on the knowledge dimension (Widiana et al., 2023), offering a structured approach to the cognitive modeling of students. Another important framework, Marzano's Taxonomy (Marzano & Kendall, 2006; Sun et al., 2023) enhances the cognitive understanding of student learning through six categories: Retrieval, Comprehension, Analysis, Knowledge Utilization, Metacognitive, and Self-System. These are distributed across three knowledge domains: Information, Mental Procedures, and Psychomotor Procedures. Marzano's framework provides more granularity in identifying and modeling student learning behaviors, further enriching the model.

The Pedagogical Relational Teachership Taxonomy (Ljungblad, 2023; Suryani & Syahbani, 2023) emphasizes the relational dynamics between teachers and students, focusing on both Tact and Stance. Each area is divided into six relational key indicators, crucial for describing and simulating the ongoing relational processes in educational settings. These indicators play a pivotal role in modeling how students engage with teachers, making them essential for creating authentic teacher–student interactions in simulations.

A key factor in modeling student profiles is their learning style. Learning styles are defined as "consistent preferences for adopting learning processes, irrespective of the task or problem presented". Learning styles could be classified along different dimensions. For example, if the dimension is learning modality, the learning styles are visual, auditory, and kinesthetic learning. If the dimension is social attitude, the learning styles are avoidant, participative, competitive, collaborative, dependent, and independent.

The Felder–Silverman Learning Styles Model (Graf et al., 2007; Ikawati et al., 2021) characterizes learners along four major dimensions, each representing a spectrum of preferences for processing and organizing information. *The first* dimension distinguishes between sensing learners, who favor concrete facts and systematic procedures, and intuitive learners, who gravitate toward abstract concepts and theoretical ideas. *The second* dimension contrasts visual learners, who rely on images, charts, and diagrams, with verbal learners, who learn effectively through spoken or written language. *The third* dimension differentiates active learners, who gain understanding by engaging in hands-on activities and discussions, from reflective learners, who require time to think through material and process information privately. Finally, *the fourth dimension* compares sequential learners, who learn in a logical, linear fashion, with global learners, who absorb information in broad strokes and often synthesize knowledge in sudden bursts of insight. Although each dimension presents dichotomous categories, individuals typically blend traits from both ends of each continuum, and recognizing these learning style differences can help educators tailor teaching methods to suit diverse learners.

Maya et al. (2021) and Clavido (2024) explored the relationship between learning styles and academic achievement, which forms a foundation for developing responsive student profiles. Hananto et al. (2024) extended this work by using Support Vector Machines (SVMs) to identify specific learning patterns in students based on Felder–Silverman learning styles. In addition to this model, the Grasha–Riechmann Learning and Teaching Styles framework (Sim & Mohd Matore, 2022) has been validated across multiple cultural contexts, including Argentina (Freiberg Hoffmann & Fernández Liporace, 2021), China (Sim & Mohd Matore, 2022), Iran (Baneshi et al., 2013), and Turkey (Baykul et al., 2010). This framework allows for modeling diverse student personas, reflecting the heterogeneity of real-world classrooms.

2.5. Personality Traits and Learning

Several studies show a strong relationship between personality and learning style (Bruso et al., 2020; Lei, 2022; Singar & Jain, 2024). Two leading personality models characterize personality through several dichotomous characteristics. The older model is the Myers–Briggs Type Indicator (MBTI), and the newer model is the Big Five personality traits. Either one could serve as a proxy for choosing a learning style.

The Myers–Briggs Type Indicator comprises the following four dimensions: Extraversion vs. Introversion, Intuition vs. Sensing, Thinking vs. Feeling, and Judgment vs. Perception. Kang and Kusuma (2020) showed that learning style is related to the MBTI score in the four dimensions. Murphy et al. (2020) found a strong relationship between MBTI personality type and preferred teaching methods for undergraduate college students. Ullah et al. (2024) explored the effect of MBTI personality type on the language learning strategies of non-English major students.

The Big Five personality traits are a group of five characteristics used to study personality:

- Openness to experience (inventive/curious vs. consistent/cautious);
- Conscientiousness (efficient/organized vs. extravagant/careless);
- Extraversion (outgoing/energetic vs. solitary/reserved);
- Agreeableness (friendly/compassionate vs. critical/judgmental);
- Neuroticism (sensitive/nervous vs. resilient/confident).

John et al. (2020) and Mammadov (2022) found a strong relationship between the Big Five personality traits and academic performance. Khalilzadeh and Khodi (2021) found a relationship between teachers' personality traits and students' motivation using structural equation modeling analysis.

2.6. Simulating Social and Emotional Complexities

Large language models (LLMs) have made notable strides in capturing emotional and social subtleties, allowing them to participate in empathetic and contextually fitting exchanges. For instance, GPT-4 can formulate replies that exhibit emotional intelligence—such as offering reassurance in stressful conversations or adjusting its tone to match a user's affect—primarily due to training on extensive datasets encompassing human dialogues, literary works, and psychological materials (Bubeck et al., 2023). These data sources enable LLMs to emulate complex emotional cues and social dynamics.

Despite their capacity to simulate human-like emotional interactions, LLMs do not possess true emotional understanding or consciousness, and their contextually appropriate responses may lack genuine empathy. One way to address this limitation is through a "human-in-the-loop" strategy, wherein human experts direct the model by specifying particular emotional or behavioral intentions. For example, a therapist integrating an LLM into mental health support could craft prompts designed to elicit compassionate and patient responses, aligning with therapeutic objectives (Fitzpatrick et al., 2017). This combined method draws on the language-processing capabilities of LLMs while ensuring that emotional and social sensitivities are shaped by human expertise. Until LLMs can autonomously replicate authentic emotional behavior, such a collaborative framework offers a viable path toward more human-like engagement in settings that require empathy and nuanced understanding.

2.7. Teacher Sense of Efficacy

The Teacher Sense of Efficacy Scale (TSES) (Duffin et al., 2012) has consistently been shown to represent three areas of teaching: efficacy for classroom management (CM), efficacy to promote student engagement (SE), and efficacy in using instructional strategies (IS). Monteiro and Forlin (2023) have shown that the long TSES with 24 items is as valid as the shorter TSES version with 12 items. An important measure of teachers' sense of efficacy is the Teaching and Learning International Survey (TALIS) (Suhandoko et al., 2024). TALIS is an annual international survey on teaching and learning worldwide. It covers teaching-related issues, including training effectiveness and teachers' sense of efficacy (An et al., 2021). For this paper, the important findings are related to the relationship between teacher training and an increased teacher's sense of efficacy. Göker (2020) found a strong relationship between cognitive teacher training and an increased teacher's sense of efficacy. Samuelsson et al. (2021) showed that simulation training gives teachers the same sense of efficacy as teaching pupils. These findings were repeatedly validated in various countries (Monteiro & Forlin, 2023).

2.8. LLM-Based Simulation Platforms for Teacher Training

There are several advanced platforms for developing multiagent applications based on large language models (LLMs), including Autogen (Wu et al., 2023) and Crew AI (CrewAI platform) described by (Berti et al., 2024). Wang et al. (2023) successfully used communication between LLM agents for agent learning. Utilizing these existing tools will allow us to focus on characterizing the system's core features.

Ruiz Rojas et al. (2022) proposed 4PADAFE, a seven-phase methodology for developing an online course while integrating AI. In Ruiz-Rojas et al. (2023), this methodology was expressed in a matrix form. Noroozi et al. (2024) highlighted pedagogical, theoretical, and methodological perspectives on using generative AI in education. Y. C. Chen and Hou (2024) described a framework for employee ethics training using ChatGPT for contextual scaffolding in an interactive educational game designed for teaching ethics. Yue et al. (2024) presented MATHVC, the very first LLM-powered virtual classroom containing multiple LLM-simulated student characters, with whom a human student can practice their math skills.

3. The Framework

3.1. Conceptual Overview of the Platform

This section describes the proposed AI-based platform for deliberate teaching practice. The diagram in Figure 1 presents a high-level description of the proposed platform.

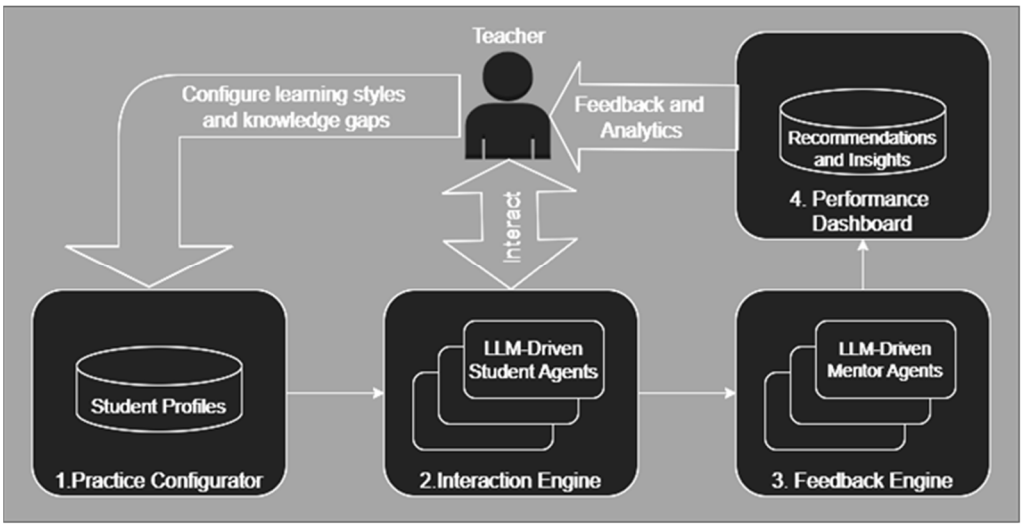


Figure 1. Proposed platform—high-level architecture.

In the proposed model (Figure 1), a teacher starts by defining the objectives and the configuration of the practice exercise (teaching subject, student behavioral patterns, types of questions asked, or expected learning profiles) via the Practice Configurator module (1). The Practice Configurator generates a definition for student agents hosted by the Interaction Engine (2). The teacher interacts with the class (via chat interface addressing an entire class or specific students). All student agents (2) are exposed to class interactions and generate responses or questions based on their individual configuration. The entire class interaction history is saved for future analysis by the mentor agents hosted by the Feedback Engine (3). Using the retained class interaction history, the mentor agents produce the teaching session evaluation and recommendation report that is submitted to the teacher via the Performance Dashboard (4).

3.2. Adaptive LLM-Driven Student Agents

The platform's LLM-driven student agents are designed to emulate diverse, real-world classroom dynamics by simulating various learning styles and behavioral characteristics based on established research. Drawing from the general personality traits and learning styles discussed in the literature review, the agents are modeled to reflect a wide range of learning preferences and personality types. This enables the simulation to offer more realistic and effective student–teacher interactions, allowing teachers to experience and respond to diverse classroom scenarios.

These agents are not static; they adapt over time, simulating student development by exhibiting growth, regression, or behavioral changes based on the teacher's instructional methods. By incorporating these dynamic elements, the platform allows teachers to refine their teaching strategies in an evolving environment, ensuring that they are better equipped to handle the complexities of real-world classrooms. The adaptive behavior of the agents is aligned with theories of learning and personality development, making the simulation a valuable tool for practicing differentiated instruction and personalized teaching.

3.3. Multi-Dimensional Mentor Agents for Feedback

The platform's mentor agents extend beyond traditional feedback systems by offering multi-dimensional insights encompassing various teaching competencies. Rather than focusing solely on lesson clarity or student engagement, these agents provide real-time feedback on critical areas such as classroom management, inclusivity, and emotional intelligence, drawing from theories like the Pedagogical Relational Teachership Taxonomy discussed in the literature review. This ensures that teachers receive comprehensive guidance on managing relational dynamics and fostering a positive classroom environment.

Each mentor agent is programmed to monitor specific aspects of teaching and classroom interactions, delivering tailored feedback that helps the teacher address areas that need improvement. Over time, the mentor agents "learn" from previous sessions, adapting their feedback to target persistent challenges, thus creating a personalized, evolving feedback loop. This iterative process supports the teacher's ongoing development, enhancing their ability to refine instructional techniques and gain deeper insights into effective classroom management and student engagement, aligned with the teacher's sense of efficacy principles.

3.4. Dynamic Classroom Simulation

The platform's classroom simulations are designed to replicate the unpredictable nature of real teaching environments, providing a far more dynamic and authentic experience than static, predefined scenarios. By leveraging insights from LLM-based simulation platforms, such as those discussed in the literature review, the platform introduces real-time adaptive simulations where student behaviors evolve based on teacher interactions. These adaptive elements simulate real-world classroom challenges like spontaneous disruptions, off-topic questions, and varying levels of student engagement.

In the proposed LLM-based class simulations, real teachers interact with a group of AI agents that form the class. These agents are AI-simulated students, having different personas and learning characteristics. It is crucial to note that all classroom communications are visible to both teachers and student agents. In our proposed scheme, the LLM agents generate student responses not only based on the teacher's prompts but also by considering interactions with other simulated student agents. This creates a dynamic and context-aware learning environment where the broader classroom discussion shapes each response. Therefore, teachers should approach these simulations with an awareness of the entire class dynamics, as previous exchanges and evolving discussions influence student responses.

This holistic perspective allows teachers to practice and refine their instructional strategies in a more immersive and interactive setting.

Teachers must adjust their instructional strategies on the fly, cultivating agility and adaptability in their teaching methods. The evolving nature of classroom simulation ensures that educators are better prepared for the complexities of managing diverse student behaviors. It provides valuable practice in navigating actual classrooms' fast-paced and often unpredictable dynamics. This real-time adaptability enhances the realism of the simulation, helping teachers develop the necessary skills to respond effectively to an evolving student body.

Large language models (LLMs) can be strategically harnessed to simulate student behavior tailored to particular local contexts or target audiences (Gao et al., 2024) through two main methods: prompt conditioning and fine-tuning.

First, *prompt conditioning* involves supplying the LLM with locally relevant information from human supervisors, ensuring that the model reflects the unique cultural, linguistic, and educational attributes of a given region or institution (Kienzler et al., 2023). For example, a supervisor might provide local idioms, historical context, or curriculum guidelines, enabling the LLM to produce responses that resonate more strongly with the target audience.

Second, *fine-tuning* on real classroom transcripts from target schools or institutions equips the LLM with a deeper understanding of the specific communication styles, academic challenges, and interaction patterns found in those environments (Ilagan et al., 2024). This process ensures that the simulated students exhibit realistic academic behaviors and align with the institution's pedagogical aims (Radford et al., 2021).

By combining these strategies, educators can create highly context-aware simulated students for teacher training, curriculum development, and educational research. This approach merges human expertise with the scalability of LLMs, allowing for customized learning experiences that meet the distinct needs of learners while preserving essential cultural and contextual nuances.

3.5. Multi-Method Adaptive Deliberate Practice

At the heart of the platform is a multi-method deliberate practice framework, where teachers engage in repeated teaching cycles of the same subject matter but with varying configurations of student agents. This approach draws from established simulation-based practice methods discussed in the literature review, allowing teachers to experiment with different instructional strategies, such as inquiry-based and project-based learning.

Each practice cycle allows teachers to focus on refining specific teaching techniques by experiencing diverse student reactions and classroom dynamics. The structured repetition ensures that teachers can analyze how different approaches affect student outcomes, helping them develop a versatile toolkit to address various classroom challenges. Throughout these cycles, mentor agents provide continuous, targeted feedback, guiding teachers in refining their methods and reinforcing key teaching strategies over time. This iterative process aligns with the principles of deliberate practice, ensuring that teachers can iteratively refine their skills in a controlled yet dynamic environment.

3.6. Evaluation Methods

Our platform employs a multi-faceted evaluation model, integrating three data-driven methods for assessing system performance: (a) subjective (Ngereja et al., 2020), (b) objective (Wen & Ji, 2020), and (c) simulation-based (Robinson et al., 2024). The subjective evaluation collects participant feedback (teachers, mentors, and students) through surveys to assess the perceived effectiveness of the training, following frameworks such as those outlined

by Fernandes et al. (2023). This method provides valuable insights into the participants' experiences but is inherently subjective.

For a more concrete analysis, we use objective evaluation, which measures student progress in a specific subject depending on the method of instruction (Grădinaru et al., 2021; Mohan, 2023). This approach evaluates the real impact of teaching strategies on student learning outcomes. Finally, simulation-based evaluation tracks improvements in student agents' scores on standardized tests, providing a reliable way to assess how effectively teachers apply instructional techniques in the simulated environment.

The SLOTE metric (Levin et al., 2023) adds a layer of comprehensive evaluation, combining quantitative and qualitative measures of simulation-based learning (SBL) effectiveness. This metric, which we adopt as a core tool in our evaluation system, is designed to measure the efficiency of teacher training simulations and the learning outcomes that they produce in field experiments.

Early examples of using AI-driven virtual students, such as the one-on-one interaction models based on ChatGPT (Markel et al., 2023), have demonstrated the potential for AI to provide meaningful feedback in simulated learning environments. Similarly, studies such as that by Smith and Holmes (2020) on deliberate practice in physics lab teaching have shown the benefits of combining AI agents with human learners.

3.7. System Architecture

The generative AI-based teacher training platform's system architecture is modular, scalable, and designed to provide personalized, adaptive simulations. Key components work together to create immersive teaching environments, offering real-time feedback and data-driven insights to enhance teacher development. The architecture is focused on ensuring that the platform is adaptable for diverse educational settings, supporting the varied needs of both novice and experienced teachers.

3.7.1. Practice Configurator

The Practice Configurator allows teachers to define parameters for each training session, including subject matter, student behavioral profiles, and class complexity. This feature personalizes simulations, enabling teachers to practice in various environments. Advanced options like adjusting students' socio-emotional characteristics enhance the configurator's flexibility.

3.7.2. Interaction Engine

The Interaction Engine manages real-time teacher–student communication through chat or voice commands, with student agents responding dynamically based on LLM-driven profiles. The engine supports both class-wide and one-on-one interactions, simulating varied teaching scenarios. Incorporating multi-modal interactions (e.g., non-verbal cues) adds depth to the simulations, providing teachers with richer experiences.

3.7.3. Feedback Engine

The Feedback Engine delivers real-time suggestions from mentor agents during simulations, analyzing teacher performance and classroom dynamics. Post-session reports highlight strengths and areas for improvement. Adaptive feedback loops tailor advice based on the teacher's progress and ensure that feedback evolves as teachers refine their skills.

3.7.4. Performance Dashboard

The Performance Dashboard aggregates key metrics, such as student performance, interaction quality, and session progress, allowing teachers to track their development

over time. Adding predictive analytics can offer targeted suggestions for future practice, enhancing personalized learning paths.

3.7.5. Integration of Real-World Data

The platform allows the integration of real-world classroom data to create simulations that reflect actual teaching challenges. These data can include student demographics, attendance, or past performance, enabling more realistic scenarios.

3.7.6. Scalability and Accessibility

The architecture is designed for scalability and accessibility, ensuring compatibility across different educational settings. Cloud-based infrastructure and multi-language support make the platform adaptable for diverse regions, while offline functionality ensures usability in areas with limited connectivity.

4. Use Case

We present a use case scenario to demonstrate the capabilities and practical application of the proposed AI-based platform. This example simulates a teaching session for a history lesson, focusing on examples of classroom interaction, student dynamics, and mentors' feedback. The goal is to illustrate how the platform supports teachers in honing their skills by interacting with virtual students and receiving tailored, real-time feedback from AI-driven mentor agents. In this use case, a history teacher is tasked with delivering a lesson on the causes of World War I. The teacher aims to engage the class in discussing key historical events while navigating student questions and maintaining classroom engagement.

4.1. Student Agents: Simulating Realistic Classroom Dynamics

In this use-case, the teacher interacts with four distinct LLM-driven student agents, each representing a different personality type, learning difficulty, and prior knowledge (Table 3). These agents are designed to simulate the diverse challenges teachers face in real classrooms based on predefined student profiles. The number of virtual students could be easily changed. Default personality features, age, gender, and other traits are randomly chosen for each student, and could be changed by the user by invoking the configuration interface.

Table 3. Examples of virtual students' profiles (personas).

Name	Age	Grade	Behavior in Class	Learning Challenges/Strengths
Emily Chen	17	**11**	Participates quickly in class discussions	Sometimes struggles to stay focused on one topic
Michael Johnson	16	**11**	Often engages with follow-up questions to clarify understanding	Finds it challenging to retain specific facts
Sophia Rogers	16	**11**	Detail-oriented, focused on task completion	Requires more time to process and answer complex questions
David Rodriguez	17	**11**	Shows strong engagement in analytical thinking	Faces challenges understanding complicated language and historical terms ESL student).

These virtual students respond to the teacher's prompts, ask clarifying questions, and showcase learning difficulties, providing the teacher with diverse classroom scenarios. The simulation allows the teacher to practice adjusting their teaching methods based on individual student needs, repeating lessons if necessary, and refining explanations for better comprehension.

4.2. Class Interaction and Engagement

The lesson begins with the teacher posing a question about the causes of World War I (Figure 2). Emily Chen, who tends to participate quickly in class discussions, immediately responds by mentioning the assassination of Archduke Franz Ferdinand. While her response is accurate, it lacks the depth and additional context that could help deepen the class discussion The teacher recognizes Emily's tendency to jump in quickly and praises her for identifying the event but moves the discussion forward to maintain her engagement and prevent her from losing focus.

Teacher: Good morning, class! Today, we'll be discussing the causes of World War I. Who can name one of the significant factors that led to the war?

Emily Chen: The assassination of Archduke Franz Ferdinand of Austria-Hungary in 1914.

Teacher: Excellent, Emily! That's often considered the immediate trigger for the war. Can anyone else name another factor?

Michael Johnson: Wasn't there also tension between European countries due to imperialism and militarism?

Teacher: Great point, Michael. Imperialism and militarism were indeed major contributing factors. Let's discuss how these elements created a volatile situation in Europe.

Figure 2. Sample of class interaction.

Michael Johnson, who often struggles with retaining specific facts, follows up with a broader observation about imperialism and militarism contributing to tensions in Europe. Despite his challenge with memory recall, Michael often engages with clarifying questions to ensure he understands the material. In this case, he asks for more details to connect the concepts of imperialism and militarism to the causes of World War I. The teacher, recognizing Michael's need for repetition, reinforces the key points by restating the events in a clear, structured manner. This repetition helps Michael solidify his understanding and aids his retention of the details. The above interaction is summarized in Figures 2 and 3.

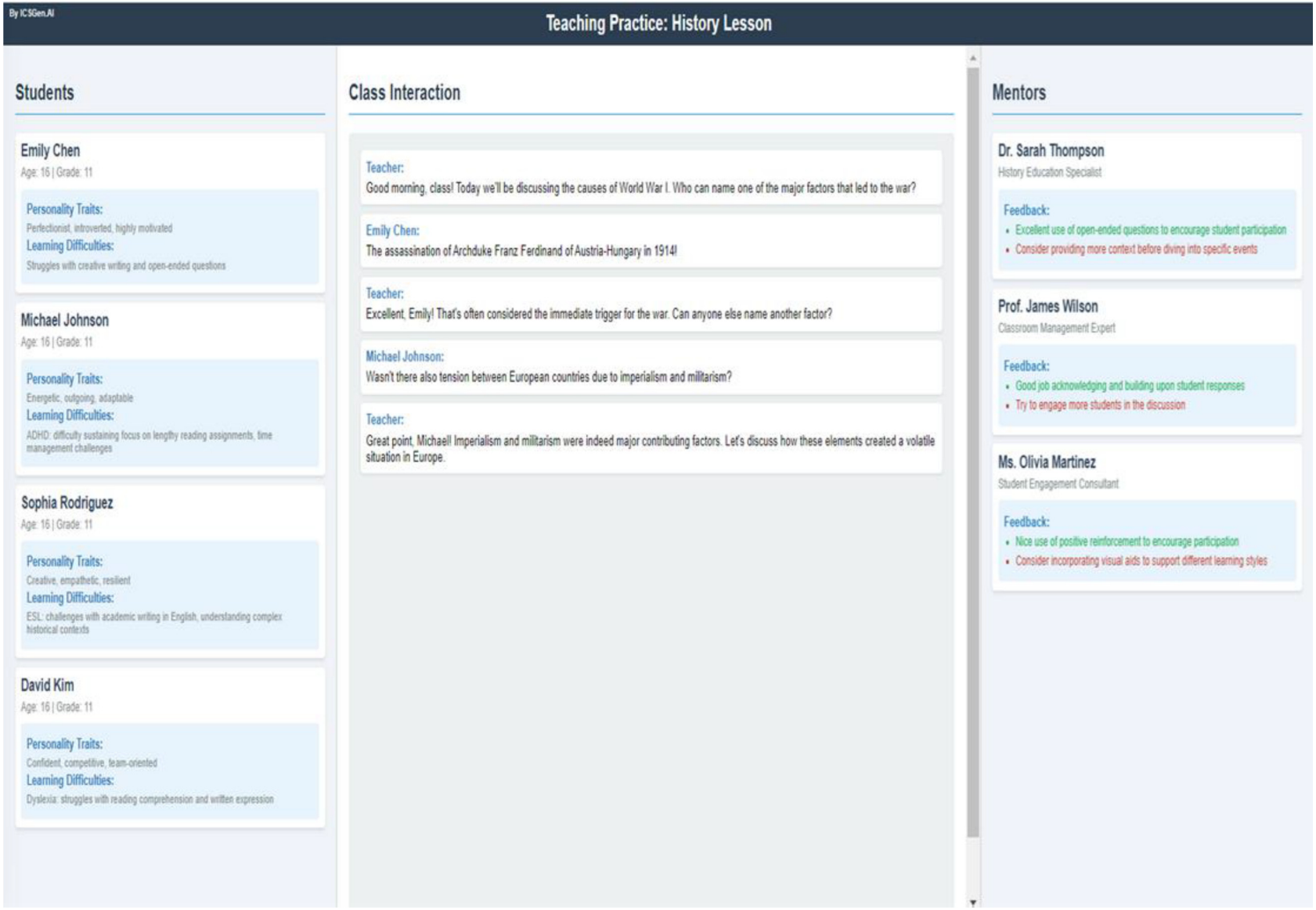


Figure 3. History teaching practice simulation example demonstrating the proposed virtual class.

The screen in Figure 3 is divided into three primary regions, each serving a distinct function. The left region of the screen displays a list of virtual students, randomly and arbitrarily developed by the generative system. The students are presented in boxes, which include randomly chosen names like "Emily Chen", "Michael Johnson", and so on. Each student was randomly assigned visible attributes, such as age, grade level, and personality traits, with brief descriptions under their names. The central section consists of a large area showing the ongoing dialogue between the teacher and the students, represented as a chat or transcript. Here, the teacher's prompts to the virtual students and their responses are shown in real time, along with the teacher's follow-up prompts and questions. On the right side of the screen are three virtual mentors. These are generated agents labeled as individuals with specialized areas of expertise, such as Dr. Sarah Thompson or Prof. James Wilson. Here, the mentors provide feedback on the teacher's actions, using text-based comments and suggestions that appear beside their names. This layout allows the user to view the interactions among the teacher, students, and mentors simultaneously.

4.3. Mentor Agents: Providing Real-Time Feedback

Throughout the use case lesson, three AI-driven mentor agents observe the class interaction and offer real-time feedback to the teacher. The mentor information and feedback are summarized in Table 4. For example, after the teacher's interaction with Michael Johnson, Dr. Sarah Thompson suggests emphasizing key historical events more explicitly to aid in memory retention. Prof. James Wilson recommends asking more open-ended questions to deepen classroom discussion. At the same time, Ms. Olivia Martínez suggests using more straightforward, more direct language to engage students like David more effectively.

Table 4. Mentors' profiles and feedback.

Mentor's Name	Expertise	Feedback	
Dr. Sarah Thompson	History Education Specialist	Strengths	Excellent use of open-ended questions to encourage student participation
		Suggestions	Consider providing more context before diving into specific events
Prof. James Wilson	Pedagogical Expert	Strengths	Good job acknowledging and building upon students' responses
		Suggestions	Try to engage more students in the discussion
Ms. Olivia Martínez	Student Engagement Consultant	Strengths	Excellent use of positive reinforcement to encourage participation
		Suggestions	Consider incorporating visual aids to support different learning styles

4.4. Teaching Performance Evaluation and Feedback

After the simulated lesson, the platform generates a performance summary based on the student agents' interaction logs and the feedback from mentor agents. This summary highlights areas of strength, such as the teacher's ability to simplify complex topics for students with language difficulties, and areas for improvement, such as providing clearer instructions to engage more reserved students. Clearly, using virtual students prevents the possibility of measuring the learning achievements of real students. So, if rigorous validation is sought, some observations on the achievements of real students are necessary. However, only a small sample may suffice for statistical validation.

Measuring teacher performance can involve analyzing class interaction transcripts through large language models (LLMs) and natural language processing (NLP) to extract quantifiable features that serve as proxies or indicators of teacher efficacy. Although these features are not exhaustive, as true efficacy encompasses more than can be inferred from discourse alone, they offer valuable insights into classroom dynamics. One dimension of analysis is the teacher–student talk ratio, in which the percentage of teacher talk relative to student talk can reveal the balance between lecturing and interactive dialogue. Turn-taking analysis further refines this assessment by determining whether a single speaker, usually the teacher, dominates the conversation or whether verbal contributions are more evenly distributed among participants.

Questioning techniques, a crucial element in effective instruction, can be gauged by tracking the frequency and types of questions posed. Open-ended, reflective, or evaluative questions generally promote deeper thinking compared to yes/no or recall-based queries. The average wait time after posing a question offers additional insight; a longer pause can encourage more meaningful student responses. Feedback patterns also influence the classroom climate and learning outcomes. The ratio of positive to corrective feedback sheds light on how often teachers provide encouragement and validation relative to criticism. The depth of feedback—whether it is specific, constructive, and linked to lesson goals or merely cursory—offers a further indicator of the teacher's ability to guide and support student learning.

Instructional clarity and structure can be investigated by measuring discourse coherence, which reflects how well the teacher organizes ideas, transitions between concepts, and maintains topical consistency. High coherence often indicates effective instructional methods. A related aspect is the use of scaffolding: markers of teacher scaffolding—such as guiding cues, prompts, or leading questions—demonstrate a commitment to building on student responses and facilitating deeper understanding. Student engagement indicators, for example, the proportion of student-initiated talk, serve as a gauge of how comfortable and motivated students feel to participate without being prompted. Additionally, the sentiment analysis of student contributions can highlight whether students exhibit a generally positive or engaged emotional tone, a factor that may correlate with a supportive classroom environment.

Vocabulary and complexity, in terms of the teacher's lexical diversity, can signal expertise and the capacity to contextualize course content. Aligning language with accurate, domain-specific terminology can be especially critical in subjects like mathematics or the natural sciences. Another domain of interest is response sensitivity and adaptation, demonstrated when teachers modify their explanations or questions based on student statements, suggesting responsiveness and an adaptive teaching style. Intertextual references, such as referring back to earlier lessons or linking new content to prior student contributions, can further reflect attentiveness to student input and continuity in instruction.

Emotional and motivational support, measured by the frequency of encouraging language or empathetic remarks, provides an indication of how well a teacher fosters a positive learning atmosphere. Personalizing feedback by using students' names or tailoring guidance to individual needs can reinforce this supportive environment. Additionally, time on task and lesson flow can be assessed by examining how smoothly the teacher transitions between activities, as well as how quickly off-topic discussions are redirected to instructional goals.

Finally, reflective talk represents a teacher's self-awareness and adaptability. Instances where the teacher explicitly explains their thought process or revises instructional methods in real time—signaled by statements such as "Let's approach this another way…"—can indicate a high level of pedagogical awareness and continuous professional growth. By analyzing these various dimensions in class interaction transcripts, researchers and practitioners can develop a more nuanced understanding of teacher performance, even though the broader concept of teacher efficacy necessarily extends beyond any single set of measurable linguistic features.

LLMs and NLP can support these metrics in several ways. First, advanced speech-to-text and segmentation models automatically transcribe spoken language and divide transcripts into speaker turns, facilitating the calculation of talk ratios and turn-taking patterns. Second, NLP models classify questions according to frameworks like Bloom's Taxonomy, distinguishing, for example, between open-ended and closed questions. Third, sentiment and emotion analysis—powered by LLM-based classifiers—gauges the tone of teacher feedback and student responses, providing insight into the classroom climate. Fourth, topic modeling and discourse analysis identify the main themes or topics under discussion, track how long each topic persists, and evaluate the coherence of the teacher's instruction. Fifth, named-entity recognition (NER) detects references to student names, academic terms, or personalizing language, serving as indicators of student-centered instruction. Lastly, by comparing initial teacher statements to subsequent modifications, NLP can quantify the teacher's responsiveness to student input or misconceptions, thus offering a measure of adaptive teaching. Figure 4 shows a sample teacher efficacy report based on the above descriptions.

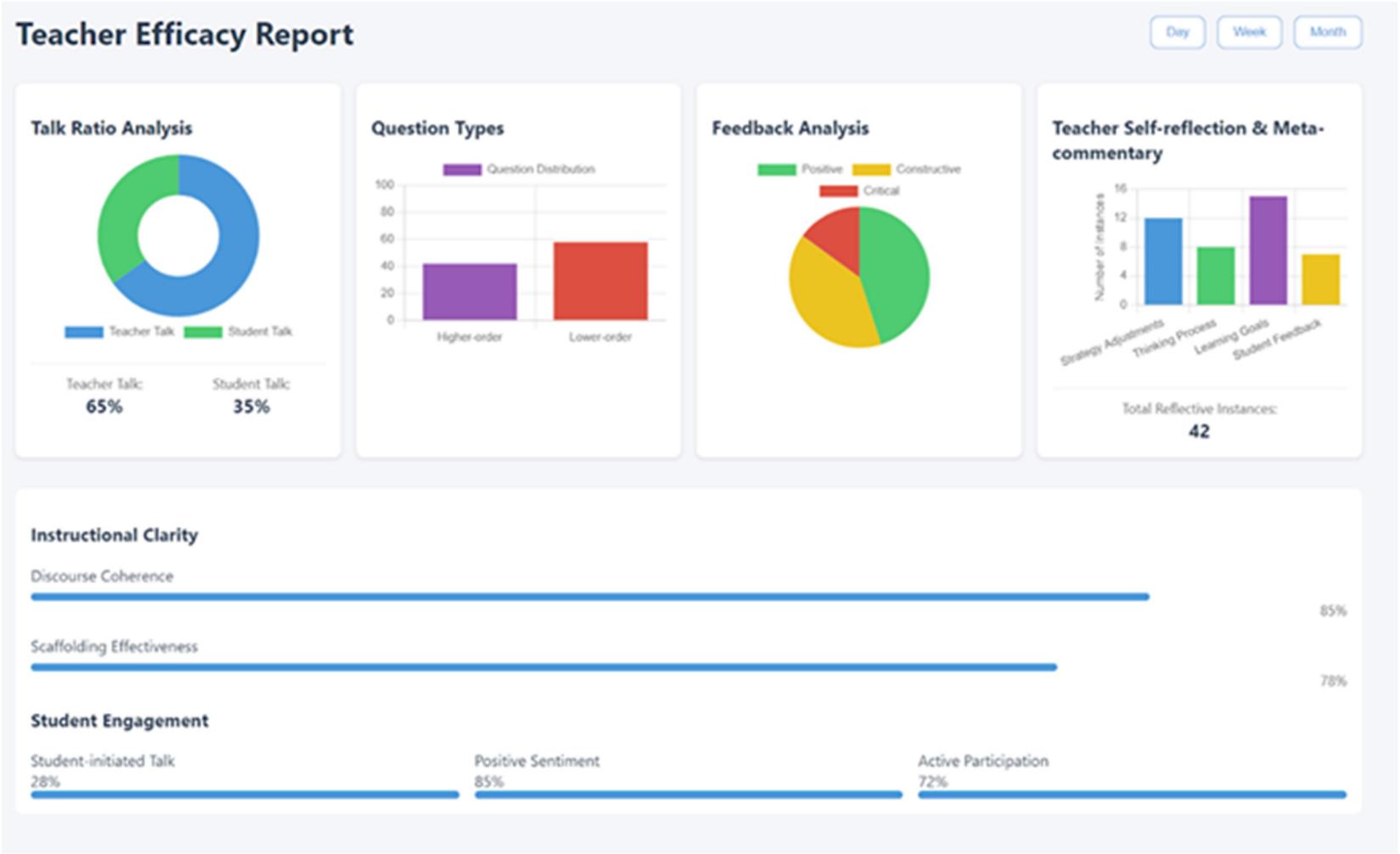


Figure 4. Teacher efficacy report.

5. Discussion

The proposed AI-driven teacher training framework offers a novel approach addressing the challenges of scalability, personalization, and adaptability in teacher education (in comparable teacher training platforms). This platform uses large language models (LLMs) and multiagent systems to provide a flexible solution for developing teaching competencies through deliberate, repeated practice. The system creates realistic, dynamic classroom simulations, allowing teachers to interact with adaptive virtual student agents and receive real-time feedback from AI-powered mentor agents.

One of the framework's key advantages is its ability to simulate diverse classroom dynamics. The platform's LLM-driven student agents can be customized with varying learning styles, cognitive abilities, and behavioral traits. These student agents offer teachers a highly realistic and flexible environment. These agents adapt based on the teacher's instructional methods, evolving to reflect student growth, regression, and various individual needs. This adaptability mirrors real-world challenges where teachers must adjust their strategies in response to diverse classroom conditions, providing a valuable practice ground for developing classroom management and differentiated instruction skills.

In the use case, the history lesson on World War I demonstrates the framework's ability to offer a rich, varied teaching experience. Here, the teacher interacts with student agents representing different learning profiles, such as language barriers and attention difficulties. The platform's adaptability is further exemplified when the teacher tailors their approach based on real-time student feedback—adjusting explanations and pacing to suit individual student needs. Mentor agents provide multi-dimensional feedback on these interactions, offering guidance not only on lesson clarity and engagement but also on deeper aspects such as managing classroom dynamics and responding to student challenges.

The platform's use of real-time, personalized feedback is a notable advantage over traditional teacher training methods. Mentor agents continuously assess the teacher's performance, delivering insights that are tailored to specific teaching scenarios. This iterative cycle of practice, feedback, and improvement is essential for developing effective teaching strategies and enhancing pedagogical skills. By focusing on both instructional techniques and classroom management, the platform supports comprehensive teacher development in a controlled environment that minimizes risks for actual students.

While the framework presents significant advantages, it also faces challenges that must be addressed to maximize its potential fully. One key issue is ensuring the realism of virtual student agents. Although the platform simulates diverse learning needs, it may struggle to replicate the emotional, cultural, and socio-economic complexities that exist in real classrooms. Enhancing the emotional intelligence of virtual agents would improve the authenticity of the simulations, allowing teachers to better practice while addressing the nuanced, interpersonal aspects of teaching.

Another concern is the quality and depth of AI-generated feedback. While the AI system provides data-driven insights quickly, it risks focusing too much on quantifiable elements like pacing and content delivery, potentially overlooking more subtle classroom dynamics such as student engagement and relational interactions. Ensuring that AI feedback complements rather than replacing human mentorship is crucial for fostering reflective, holistic teaching practices.

Additionally, ethical considerations regarding AI bias and data privacy are paramount. Biases in AI-driven feedback or student simulations could inadvertently reinforce harmful stereotypes, affecting both teacher and student development. Addressing these issues will require robust ethical frameworks to guide the responsible deployment of AI in education, ensuring that the system promotes fairness, transparency, and inclusivity.

Since technology mediates how people interact with the world around them and understand it, the very use of technology by itself is a factor that influences the beliefs and values, and hence the actions, of the human agents involved. Therefore, we advocate that using simulation for teacher training should be accompanied by training on real classes with real human students. This would help mitigate the biases in beliefs and values of beginning teachers who have been developed through 10,000 h of generative AI simulation-based training. However, we still are in the dark as to how technology-based training will affect the teachers in the classrooms they are assigned to teach. How would this change how classrooms are managed and how students are taught, and what further downstream impacts would this have on future generations?

Despite these challenges, the framework's advantages are substantial. It provides a scalable, accessible platform that supports teachers in refining their skills through deliberate practice in a low-risk, controlled environment. The combination of real-time AI feedback with adaptive simulations offers a level of personalization and responsiveness that traditional training methods cannot achieve. By enabling teachers to engage with diverse, evolving classroom scenarios, the platform fosters the development of critical competencies needed to succeed in today's complex educational environments.

6. Conclusions

The proposed AI-driven teacher training framework offers a scalable and adaptive solution for enhancing teacher education through personalized simulations and utilizing large language models (LLMs) and multiagent systems. The platform simulates diverse classroom environments, enabling teachers to practice with virtual students and receive real-time feedback from AI-driven mentor agents. This approach addresses key challenges in traditional teacher training, such as scalability and the need for repeated, personalized practice and continuous feedback.

The framework supports deliberate practice, with adaptive student agents created using broad theoretical foundations of various learning styles and personalities. The student agents evolve based on teacher interactions, helping educators manage diverse student needs. Mentor agents provide multi-dimensional feedback, allowing teachers to refine their instructional strategies and classroom management skills. A use case demonstrates the platform's flexibility in simulating various teaching scenarios, promoting the development of differentiated instruction and classroom engagement. By aligning these capabilities with the study's objectives, this framework directly contributes to addressing the need for scalable, adaptive, and feedback-driven teacher training. The results indicate that AI-driven simulations provide a structured, iterative environment where teachers can improve instructional techniques while adapting to diverse classroom dynamics. Additionally, the incorporation of mentor agents facilitates real-time assessment, ensuring that teachers receive actionable insights to refine their pedagogical approaches. These findings validate the potential of generative AI to enhance professional development in education, bridging the gap between traditional training methods and the evolving demands of modern classrooms.

However, challenges remain, such as ensuring the realism of virtual agents, maintaining the quality of AI feedback, and addressing ethical concerns like data privacy and AI bias. While this study establishes a conceptual foundation, further validation through empirical classroom studies will be essential to fully assess the effectiveness of AI-driven teacher training. Future implementations should explore how real-world educators interact with the platform, measuring improvements in instructional efficacy, engagement, and learning outcomes. Despite these limitations, the platform can potentially transform teacher training by offering scalable, data-driven insights that complement human mentorship.

The development of AI-based simulation platforms for teacher education intersects with core ethical and philosophical issues in pedagogy, raising questions about the nature of teaching, learning, and the role of technology in education. One fundamental dilemma is the tension between standardization and individuality in pedagogy. While AI platforms can provide consistent, scalable training experiences, they risk oversimplifying the complex, context-dependent nature of teaching, which often requires adaptability and empathy (Bartholomay, 2022; Dunn & Larson, 2023). This aligns with the philosophical debate over whether education should focus on measurable outcomes or the holistic development of students, a concern highlighted by Miseliunaite et al. (2022) in their critique of technocratic approaches to education. Additionally, using AI in teacher training raises ethical concerns about data privacy and surveillance, as fine-tuning LLMs on class transcripts or student interactions may inadvertently expose sensitive information (Selwyn, 2019).

Furthermore, the reliance on AI simulations may devalue the irreplaceable human elements of teaching, such as moral guidance and emotional connection, which are central to the pedagogical philosophy of thinkers like Freire (2020), who emphasized the importance of dialogue and critical consciousness in education. To address these issues, AI-based platforms must be designed with ethical frameworks that prioritize transparency, inclusivity, and respect for the human dimensions of teaching. By engaging with these philosophical and moral dilemmas, educators and developers can ensure that AI tools complement, rather than undermine, the core values of pedagogy.

Future research should focus on assessing the long-term impact of AI-driven training on teacher performance, improving the emotional intelligence of AI agents, and developing ethical frameworks to guide the responsible use of AI in education. Empirical studies should explore how teachers engage with AI-driven simulations in practice, evaluating both skill acquisition and transferability to real classroom settings. Additionally, interdisciplinary collaboration between AI researchers, education experts, and policymakers will be crucial in refining the framework for widespread implementation. This framework could significantly enhance the quality and accessibility of teacher training. It provides scalability and satisfies the need for repeated, personalized practice and continuous feedback. The proposed framework ensures that educators are better prepared to meet the complexities of real-world classrooms.

Author Contributions: All authors contributed to the study conception and design. However, major contributions to the following components are as follows: Conceptualization: Y.A.; Methodology: Y.A. and Y.C.; Formal analysis and investigation: Y.C. and A.A.; Writing—original draft preparation: Y.A. and Y.C.; Writing—review and editing: A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data is contained within the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
GPT	Generative Pretrained Transformer
LLM	Large Language Model
MBTI	Myers–Briggs Type Indicator
TALIS	Teaching and Learning International Survey

References

Abedi, E. A. (2023). Tensions between technology integration practices of teachers and ICT in education policy expectations: Implications for change in teacher knowledge, beliefs and teaching practices. *Journal of Computers in Education*, *11*(4), 1215–1234. [CrossRef]

Agyei, E. D. (2024). Pedagogical support structures for effective implementation of simulation-based innovation in science classrooms: Prospective teachers' perspectives. *African Journal of Educational Studies in Mathematics and Sciences*, *20*(1), 47–52.

Alsafari, B., Atwell, E., Walker, A., & Callaghan, M. (2024). Towards effective teaching assistants: From intent-based chatbots to LLM-powered teaching assistants. *Natural Language Processing Journal*, *8*, 100101. [CrossRef]

An, Y., Li, L., & Wei, X. (2021). What influences teachers' self-efficacy in east Asia? Evidence from the 2018 teaching and learning international survey. *Social Behavior and Personality: An International Journal*, *49*(5), 1–13. [CrossRef]

Badiee, F., & Kaufman, D. (2015). Design evaluation of a simulation for teacher education. *Sage Open*, *5*(2), 2158244015592454. [CrossRef]

Baneshi, A. R., Karamdoust, N. A., & Hakimzadeh, R. (2013). Validity reliability of the Persian version of Grasha-Richmann student learning styles scale. *Journal of Advances in Medical Education & Professionalism*, *1*(4), 119–124.

Bartholomay, D. J. (2022). A time to adapt, not "return to normal": Lessons in compassion and accessibility from teaching during COVID-19. *Teaching Sociology*, *50*(1), 62–72. [CrossRef]

Baykul, Y., Gürsel, M., Sulak, H., Ertekin, E., Dülger, O., Aslan, Y., & Büyükkarcı, K. (2010). A validity and reliability study of Grasha-Riechmann student learning style scale. *World Academy of Science, Engineering and Technology*, *5*(3), 177–184.

Berti, A., Maatallah, M., Jessen, U., Sroka, M., & Ghannouchi, S. A. (2024). Re-Thinking process mining in the AI-Based agents Era. *arXiv*, arXiv:2408.07720.

Bhowmik, S., West, L., Barrett, A., Zhang, N., Dai, C. P., Sokolikj, Z., Southerland, S., Yuan, X., & Ke, F. (2024). Evaluation of an LLM-powered student agent for teacher training. In *European conference on technology enhanced learning* (pp. 68–74). Springer Nature.

Bruso, J., Stefaniak, J., & Bol, L. (2020). An examination of personality traits as a predictor of using self-regulated learning strategies and considerations for online instruction. *Educational Technology Research and Development*, *68*(5), 2659–2683. [CrossRef]

Bubeck, S., Chadrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., & Nori, H. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv*, arXiv:2303.12712.

Chandio, M. T., Zafar, N., & Solangi, G. M. (2021). Bloom's taxonomy: Reforming pedagogy through assessment. *Journal of Education and Educational Development*, *8*(1), 109–140. [CrossRef]

Chen, R. H. (2022). Effects of deliberate practice on blended learning sustainability: A community of inquiry perspective. *Sustainability*, *14*(3), 1785. [CrossRef]

Chen, Y. C., & Hou, H. T. (2024). A Mobile contextualized educational game framework with ChatGPT interactive scaffolding for employee ethics training. *Journal of Educational Computing Research*, *62*(7), 1737–1762. [CrossRef]

Clavido, G. (2024). Teaching styles and pupils' learning styles: Their relationship with pupils' academic performance. *Spring Journal of Arts, Humanities and Social Sciences*, *3*(2), 11–15. [CrossRef]

Cranford, S. (2022). Ten thoughts from a 10,000-hour "expert" editor. *Matter*, *5*(10), 3067–3071. [CrossRef]

Cuban, L. (2010). *How long does it take to become a "good" teacher?* Available online: https://larrycuban.wordpress.com/2010/04/20/how-long-does-it-take-to-become-a-good-teacher/ (accessed on 18 March 2025).

Delamarre, A., Shernoff, E., Buche, C., Frazier, S., Gabbard, J., & Lisetti, C. (2021). The interactive virtual training for teachers (IVT-T) to practice classroom behavior management. *International Journal of Human-Computer Studies*, *152*, 102646. [CrossRef]

Duffin, L. C., French, B. F., & Patrick, H. (2012). The teachers' sense of efficacy scale: Confirming the factor structure with beginning pre-service teachers. *Teaching and Teacher Education*, *28*(6), 827–834. [CrossRef]

Dunn, M. S., & Larson, K. E. (2023). Mindful educators: Compassion, community, and adaptability are key. *International Journal of Learning, Teaching and Educational Research*, *22*(2), 358–376. [CrossRef]

Eldan, R., & Russinovich, M. (2023). Who is Harry Potter? Approximate Unlearning in LLMs. *arXiv*, arXiv:2310.02238.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363. [CrossRef]

Extremera, J., Vergara, D., Gómez, A. I., Fernández, P., Ordóñez, E., & Rubio, M. P. (2020). Impediments to the development of immersive virtual reality in education. In *EDULEARN20 proceedings* (pp. 1282–1288). IATED.

Fernandes, S., Araújo, A. M., Miguel, I., & Abelha, M. (2023). Teacher professional development in higher education: The impact of pedagogical training perceived by teachers. *Education Sciences*, *13*(3), 309. [CrossRef]

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): A Randomized controlled trial. *JMIR Mental Health*, *4*(2), e19. [CrossRef]

Forehand, M. (2010). Bloom's taxonomy. *Emerging Perspectives on Learning, Teaching, and Technology*, *41*(4), 47–56.

Freiberg Hoffmann, A., & Fernández Liporace, M. (2021). Grasha–Riechmann student learning style scales: An Argentinian version. *Journal of Applied Research in Higher Education*, *13*(1), 242–257. [CrossRef]

Freire, P. (2020). Pedagogy of the oppressed. In *Toward a sociology of education* (pp. 374–386). Routledge.

Gao, C., Lan, X., Li, N., Yuan, Y., Ding, J., Zhou, Z., Xu, F., & Li, Y. (2024). Large language models empowered agent-based modeling and simulation: A survey and perspectives. *Humanities and Social Sciences Communications*, *11*(1), 1–24. [CrossRef]

Gaviria, D., Arango, J., Valencia-Arias, A., Bran-Piedrahita, L., Rojas Coronel, Á. M., & Romero Díaz, A. (2024). Simulator-mediated learning: Enhancing accounting teaching-learning processes in higher education. *Cogent Education*, *11*(1), 2340856. [CrossRef]

Gladwell, M. (2008). *Outliers: The story of success*. Hachette UK.

Göker, S. D. (2020). Cognitive coaching: A powerful supervisory tool to increase teacher sense of efficacy and shape teacher identity. *Teacher Development*, *24*(4), 559–582. [CrossRef]

Graf, S., Viola, S. R., Leo, T., & Kinshuk. (2007). In-depth analysis of the Felder-Silverman learning style dimensions. *Journal of Research on Technology in Education*, *40*(1), 79–93. [CrossRef]

Grădinaru, A. C., Spataru, M. C., & Pavel, G. (2021). The importance of objective evaluations in stimulating fair competitiveness in higher veterinary education. In *ICERI2021 proceedings* (pp. 1661–1665). IATED.

Han, Z., Gao, C., Liu, J., & Zhang, S. Q. (2024). Parameter-efficient fine-tuning for large models: A comprehensive survey. *arXiv*, arXiv:2403.14608.

Hananto, A. R., Musdholifah, A., & Wardoyo, R. (2024). Identifying student learning styles using support vector machine in felder-silverman model. *Journal of Applied Data Sciences*, *5*(3), 1495–1507. [CrossRef]

Hou, J., Ao, C., Wu, H., Kong, X., Zheng, Z., Tang, D., Li, C., Hu, X., Xu, R., Ni, S., & Yang, M. (2024). E-EVAL: A comprehensive chinese K-12 education evaluation benchmark for large language models. *arXiv*, arXiv:2401.15927.

Hu, B., Zheng, L., Zhu, J., Ding, L., Wang, Y., & Gu, X. (2024). Teaching plan generation and evaluation with GPT-4: Unleashing the potential of LLM in instructional design. *IEEE Transactions on Learning Technologies*, *17*, 1445–1459. [CrossRef]

Hu, Z., Wang, L., Lan, Y., Xu, W., Lim, E., Bing, L., Xu, X., Poria, S., & Lee, R. (2023). LLM-adapters: An adapter family for parameter-efficient fine-tuning of large language models. *arXiv*, arXiv:2304.01933.

Ikawati, Y., Al Rasyid, M. U. H., & Winarno, I. (2021). Student behavior analysis to predict learning styles based Felder Silverman model using ensemble tree method. *EMITTER International Journal of Engineering Technology*, *9*(1), 92–106. [CrossRef]

Ilagan, M., Klebanov, B. B., & Mikeska, J. (2024, June 20). *Automated evaluation of teacher encouragement of student-to-student interactions in a simulated classroom discussion*. 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024) (pp. 182–198), Mexico City, Mexico.

Ingersoll, R. M., & Strong, M. (2011). The impact of induction and mentoring programs for beginning teachers: A critical review of the research. *Review of Educational Research*, *81*(2), 201–233. [CrossRef]

Jiang, H., Zhang, X., Cao, X., Kabbara, J., & Roy, D. (2023). Personallm: Investigating the ability of GPT-3.5 to express personality traits and gender differences. *arXiv*, arXiv:2305.02547.

John, R., John, R., & Rao, Z. U. R. (2020). The Big Five personality traits and academic performance. *Journal of Law and Social Studies*, *2*(1), 10–19. [CrossRef]

Kang, H., & Kusuma, G. P. (2020). The effectiveness of personality-based gamification model for foreign vocabulary online learning. *Advances in Science, Technology and Engineering Systems*, *5*(2), 261–271. [CrossRef]

Khalilzadeh, S., & Khodi, A. (2021). Teachers' personality traits and students' motivation: A structural equation modeling analysis. *Current Psychology*, *40*(4), 1635–1650. [CrossRef]

Kienzler, J., Voss, T., & Wittwer, J. (2023). Student teachers' conceptual knowledge of operant conditioning: How can case comparison support knowledge acquisition? *Instructional Science*, *51*(4), 639–659. [CrossRef]

La Cava, L., Costa, D., & Tagarelli, A. (2024). Open models, closed minds? On agents capabilities in mimicking human personalities through open large language models. *arXiv*, arXiv:2401.07115.

Landon-Hays, M., Peterson-Ahmad, M. B., & Frazier, A. D. (2020). Learning to teach: How a simulated learning environment can connect theory to practice in general and special education educator preparation programs. *Education Sciences*, *10*(7), 184. [CrossRef]

Ledger, S., Mailizar, M., & Gregory, S. (2025). Learning to teach with simulation: Historical insights. *Journal of Compuersin. Education 12*, 339–366. [CrossRef]

Lee, U., Lee, S., Koh, J., Jeong, Y., Jung, H., Byun, G., Lee, Y., Moon, J., Lim, J., & Kim, H. (2023, December 15). *Generative agent for teacher training: Designing educational problem-solving simulations with large language model-based agents for pre-service teachers*. NeurIPS'23 Workshop on Generative AI for Education (GAIED), New Orleans, LA, USA.

Lei, J. (2022, November 26–28). *The Relationship between personality and dominant learning style*. 2021 International Conference on Education, Language, and Art (ICELA 2021) (pp. 1133–1137), Virtually.

Levin, O., Frei-Landau, R., & Goldberg, C. (2023). Development and validation of a scale to measure the simulation-based learning outcomes in teacher education. In *Frontiers in education* (Vol. 8, p. 1116626). Frontiers Media SA.

Levin, O., & Muchnik-Rozanov, Y. (2022). Professional development during simulation-based learning: Experiences and insights of preservice teachers. *Journal of Education for Teaching*, *49*(1), 120–136. [CrossRef]

Li, J., Yuan, Y., & Zhang, Z. (2024). Enhancing LLM factual accuracy with rag to counter hallucinations: A case study on domain-specific queries in private knowledge-bases. *arXiv*, arXiv:2403.10446.

Li, Y., Zhang, Y., & Sun, L. (2023). Meta agents: Simulating interactions of human behaviors for llm-based task-oriented coordination via collaborative generative agents. *arXiv*, arXiv:2310.06500.

Ljungblad, A. L. (2023). Key indicator taxonomy of relational teaching. *Journal of Education for Teaching*, *49*(5), 785–797. [CrossRef]

Lyu, H., Cheng, Y., Fu, Y., & Yang, Y. (2024, April 19–21). *Exploring A LLM-based ubiquitous learning model for elementary and middle school teachers*. 2024 6th International Conference on Computer Science and Technologies in Education (CSTE) (pp. 171–174), Xi'an, China.

Malik, R., Sharma, A., & Chaudhary, P. (Eds.). (2024). *Transforming education with virtual reality*. John Wiley & Sons.

Mammadov, S. (2022). Big Five personality traits and academic performance: A meta-analysis. *Journal of Personality*, *90*(2), 222–255. [CrossRef]

Marcus, C. H., Newman, L. R., Winn, A. S., Antanovich, K., Audi, Z., Cohen, A., Hirsch, A. W., Harris, H. K., A Miller, K., & Michelson, C. D. (2020). TEACH and repeat: Deliberate practice for teaching. *The Clinical Teacher*, *17*(6), 688–694. [CrossRef]

Markel, J. M., Opferman, S. G., Landay, J. A., & Piech, C. (2023, July 20–22). *GPTeach: Interactive TA training with GPT-based students*. Tenth ACM Conference on Learning@ Scale (pp. 226–236), Copenhagen, Denmark.

Marzano, R. J., & Kendall, J. S. (Eds.). (2006). *The new taxonomy of educational objectives*. Corwin Press.

Maya, J., Luesia, J. F., & Pérez-Padilla, J. (2021). The relationship between learning styles and academic performance: Consistency among multiple assessment methods in psychology and education students. *Sustainability*, *13*(6), 3341. [CrossRef]

Mikeska, J., Howell, H., Dieker, L., & Hynes, M. (2021). Understanding the role of simulations in K-12 mathematics and science teacher education: Outcomes from a teacher education simulation conference. *Contemporary Issues in Technology and Teacher Education*, *21*(3), 781–812.

Miseliunaite, B., Kliziene, I., & Cibulskas, G. (2022). Can holistic education solve the world's problems: A systematic literature review. *Sustainability*, *14*(15), 9737. [CrossRef]

Mohan, R. (2023). *Measurement, evaluation and assessment in education*. PHI Learning Pvt. Ltd.

Monteiro, E., & Forlin, C. (2023). Validating the use of the 24-item long version and the 12-item short version of the Teachers' Sense of Efficacy Scale (TSES) for measuring teachers' self-efficacy in Macao (SAR) for inclusive education. *Emerald Open Research*, *1*(3). [CrossRef]

Murphy, L., Eduljee, N. B., Croteau, K., & Parkman, S. (2020). Relationship between personality type and preferred teaching methods for undergraduate college students. *International Journal of Research in Education and Science*, *6*(1), 100–109. [CrossRef]

Ngereja, B., Hussein, B., & Andersen, B. (2020). Does project-based learning (PBL) promote student learning? A performance evaluation. *Education Sciences*, *10*(11), 330. [CrossRef]

Nichols, C. R., Tandstad, T., William, T., Lowrance, S., & Daneshmand, S. (2018). Ten thousand attentive hours, rapid learning, dissemination of knowledge, and the future of experience-based care in germ-cell tumors. *Annals of Oncology*, *29*(2), 289–290. [CrossRef]

Noroozi, O., Soleimani, S., Farrokhnia, M., & Banihashem, S. K. (2024). Generative AI in education: Pedagogical, theoretical, and methodological perspectives. *International Journal of Technology in Education*, *7*(3), 373–385. [CrossRef]

OpenAI. (2023). Gpt-4 technical report. *arXiv*, arXiv:2303.08774.

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023, September 28–October 1). *Generative agents: Interactive simulacra of human behavior*. 36th Annual ACM Symposium on User Interface Software and Technology (pp. 1–22), San Francisco, CA, USA.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., & Krueger, G. (2021, July 18–24). *Learning transferable visual models from natural language supervision*. 38th International Conference on Machine Learning (ICML), Virtual.

Ren, S., Cui, Z., Song, R., Wang, Z., & Hu, S. (2024). Emergence of social norms in large language model-based agent societies. *arXiv*, arXiv:2403.08251.

Robinson, S. J., Oo, Y. M., Ljuhar, D., McLeod, E., Pacilli, M., & Nataraja, R. M. (2024). A guide to outcome evaluation of simulation-based education programs in low and middle-income countries. *ANZ Journal of Surgery*, *94*, 1011–1020. [CrossRef]

Ruiz-Rojas, L. I., Acosta-Vargas, P., De-Moreta-Llovet, J., & Gonzalez-Rodriguez, M. (2023). Empowering education with generative artificial intelligence tools: Approach with an instructional design matrix. *Sustainability*, *15*(15), 11524. [CrossRef]

Ruiz Rojas, L. I., Castillo, C., & Cañizares, S. (2022). Digital teaching skills to design virtual learning classrooms with the 4PADAFE methodology. In *International conference on technological ecosystems for enhancing multiculturality* (pp. 1062–1071). Springer Nature.

Samuelsson, M., Samuelsson, J., & Thorsten, A. (2021). Simulation training is as effective as teaching pupils: Development of efficacy beliefs among pre-service teachers. *Journal of Technology and Teacher Education*, *29*(2), 225–251.

Sejnowski, T. J. (2023). Large language models and the reverse Turing test. *Neural Computation*, *35*(3), 309–342. [CrossRef] [PubMed]
Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. John Wiley & Sons.
Si, N., Zhang, H., Chang, H., Zhang, W., Qu, D., & Zhang, W. (2023). Knowledge unlearning for LLMs: Tasks, methods, and challenges. *arXiv*, arXiv:2311.15766.
Sim, S. H., & Mohd Matore, M. E. E. (2022). The relationship of Grasha–Riechmann teaching styles with teaching experience of National-Type Chinese primary schools mathematics teacher. *Frontiers in Psychology*, *13*, 1028145. [CrossRef]
Singar, A. V., & Jain, S. (2024). The role of personality traits on learning styles of engineering and management students studying the Internet of Things knowledge areas. *International Journal of Knowledge and Learning*, *17*(5), 511–528. [CrossRef]
Smith, E. M., & Holmes, N. G. (2020). Evaluating instructional labs' use of deliberate practice to teach critical thinking skills. *Physical Review Physics Education Research*, *16*(2), 020150. [CrossRef]
Soliman, M., Ali, R. A., & Khalid, J. (2024). Modelling continuous intention to use generative artificial intelligence as an educational tool among university students: Findings from PLS-SEM and ANN. *Journal of Computers in Education*. [CrossRef]
Sonnemann, J., & Blane, N. (2024). *Training new teachers with digital simulations: Evidence summary report, Analysis and Policy Observatory (APO)*. Available online: https://apo.org.au/node/326877 (accessed on 18 March 2025).
Stich, A. E., & Cipollone, K. (2021). In and through the urban educational "reform churn": The illustrative power of qualitative longitudinal research. *Urban Education*, *56*(3), 484–510. [CrossRef]
Suhandoko, A. D. J., Rahayu, U., Aisyah, S., Andayani, A., & Mikaresti, P. (2024). Teaching and Learning International Survey (TALIS) 2018: A descriptive analysis of Teacher Professional Development (TPD) in Indonesia. *Jurnal Kependidikan Islam*, *14*(2), 142–159. [CrossRef]
Sun, L., Hu, L., Zhou, D., & Yang, W. (2023). Evaluation and developmental suggestions on undergraduates' computational thinking: A theoretical framework guided by Marzano's new taxonomy. *Interactive Learning Environments*, *31*(10), 6588–6610. [CrossRef]
Suryani, N. Y., & Syahbani, M. H. (2023). Gamification and solo taxonomy: A strategy to promote active engagement and discipline in english language learning. *English Review: Journal of English Education*, *11*(3), 777–786. [CrossRef]
Törnberg, P., Valeeva, D., Uitermark, J., & Bail, C. (2023). Simulating social media using large language models to evaluate alternative news feed algorithms. *arXiv*, arXiv:2310.05984.
Ullah, A., Uddin, F., & Khan, S. (2024). Exploring the impact of MBTI personality types on teaching methods. *Qlantic Journal of Social Sciences*, *5*(3), 309–323. [CrossRef]
Urueta, S. H. (2023). Challenges facing the adoption of VR for language education: Evaluating dual-frame system design as a possible solution. *International Journal of Information and Education Technology*, *13*(6), 1001–1008. [CrossRef]
Wang, K., Lu, Y., Santacroce, M., Gong, Y., Zhang, C., & Shen, Y. (2023). Adapting LLM agents through communication. *arXiv*, arXiv:2310.01444.
Wen, F., & Ji, Z. (2020, October 20–22). *Evaluation of the learning performance for virtual simulation experiment*. 4th International Conference on Computer Science and Application Engineering (pp. 1–5), Sanya, China.
Widiana, I. W., Triyono, S., Sudirtha, I. G., Adijaya, M. A., & Wulandari, I. G. A. A. M. (2023). Bloom's revised taxonomy-oriented learning activity to improve reading interest and creative thinking skills. *Cogent Education*, *10*(2), 2221482. [CrossRef]
Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., Jiang, L., Zhang, X., Zhang, S., Liu, J., & Awadallah, A. H. (2023). Autogen: Enabling next-gen llm applications via multi-agent conversation framework. *arXiv*, arXiv:2308.08155.
Yu, J., Zhang, Z., Zhang-li, D., Tu, S., Hao, Z., Li, R. M., Li, H., Wang, Y., Li, H., Gong, L., & Cao, J. (2024). From MOOC to MAIC: Reshaping online teaching and learning through LLM-driven agents. *arXiv*, arXiv:2409.03512.
Yue, M., Mifdal, W., Zhang, Y., Suh, J., & Yao, Z. (2024). MathVC: An LLM-simulated multi-character virtual classroom for mathematics education. *arXiv*, arXiv:2404.06711.
Zhang, Y., Radishian, C., Brunswicker, S., Whitenack, D., & Linna, D. W., Jr. (2024). Empathetic language in LLMs under prompt engineering: A comparative study in the legal field. *Procedia Computer Science*, *244*, 308–317. [CrossRef]
Zhang, Z., Zhang-Li, D., Yu, J., Gong, L., Zhou, J., Hao, Z., Jiang, J., Cao, J., Liu, H., Liu, Z., & Hou, L. (2024). Simulating classroom education with LLM-empowered agents. *arXiv*, arXiv:2406.19226.
Zhao, Q., Wang, J., Zhang, Y., Jin, Y., Zhu, K., Chen, H., & Xie, X. (2023). Compete AI: Understanding the competition behaviors in large language model-based agents. *arXiv*, arXiv:2310.17512.
Zheng, P., Yang, J., Lou, J., & Wang, B. (2024). Design and application of virtual simulation teaching platform for intelligent manufacturing. *Scientific Reports*, *14*(1), 12895. [CrossRef]
Zhou, X., Zhu, H., Mathur, L., Zhang, R., Yu, H., Qi, Z., Morency, L. P., Bisk, Y., Fried, D., Neubig, G., & Sap, M. (2023). Sotopia: Interactive evaluation for social intelligence in language agents. *arXiv*, arXiv:2310.11667.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.