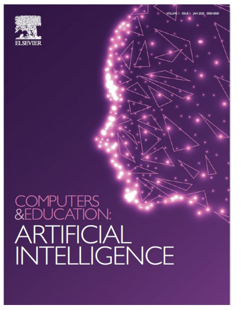
Journal Pre-proof



Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review

Chih-Pu Dai, Fengfeng Ke

PII: S2666-920X(22)00042-X

DOI: https://doi.org/10.1016/j.caeai.2022.100087

Reference: CAEAI 100087

To appear in: *Computers and Education: Artificial Intelligence*

Received Date: 30 January 2022

Revised Date: 23 June 2022

Accepted Date: 26 June 2022

Please cite this article as: Dai C.-P. & Ke F., Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review, *Computers and Education: Artificial Intelligence* (2022), doi: https://doi.org/10.1016/j.caeai.2022.100087.

Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review

Chih-Pu Dai, Fengfeng Ke

Department of Educational Psychology and Learning Systems, Florida State University, Tallahassee, USA

Acknowledgements

The work is most closely related to the National Science Foundation grant 2110777. Any opinions, findings, and conclusions or recommendations expressed in these materials are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Author Note

Correspondence should be addressed to Chih-Pu Dai, Department of Educational Psychology and Learning Systems, College of Education, Florida State University., 1114 W. Call Street, Tallahassee, FL 32306. Email: cdai@fsu.edu.

Declaration of Interest: none

Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review

Abstract

The field of education has experienced a transformation as artificial intelligence (AI) becomes increasingly applicable for learning purposes. AI has the potential to transform the social interactions in educational contexts among learners, teachers, and technologies. In this systematic mapping review, we focus on mapping and framing trends for educational applications of AI in simulation-based learning. Fifty-nine studies met the inclusion and exclusion criteria. We coded and analyzed six mapped categories in this literature review: (1) the year-of-study trend, (2) methods, (3) AI technologies, (4) simulation, (5) study trends, and (6) learning principles and theories. To provide nuanced details from the included literature, we also synthesized three thematic trends: (1) AI built in virtual agents for simulation-based learning, (2) AI infused in simulation-based learning with affective computing, and (3) AI leveraged in simulation-based learning for assessments. Trend One builds on a general acknowledgement of virtual agents as a guide for situated learning. Trend Two posits the role of affective states in learning trajectories and suggests the related machine learning approaches. Trend Three discusses machine learning techniques and multimodal computing used for assessment and feedback. The paper concludes with implications and suggestions for research and practice in AI in education using simulation-based learning.

Keywords: Artificial intelligence; simulations; virtual agents; affective computing; assessments

1. Introduction

The field of education has experienced a transformation as artificial intelligence (AI) becomes increasingly applicable for learning purposes. AI can be generally defined as "machine intelligence which is demonstrated by a non-living entity compared to natural intelligence as displayed by humans and other living species" (Leahy et al., 2019, p. 6). Artificial intelligence in education (AIEd) is a relatively young field (Chen et al., 2020) and the applications of AIEd have fallen behind other fields such as applied science and finance (Luckin & Cukurova, 2019). Despite falling behind, Chen et al. (2020) suggested that there are continued interests and potential positive impacts of AIEd research. Emerging applications of AIEd have been implemented in different educational settings, such as teachers' orchestration supports in classrooms (Holstein et al., 2019), learning analytics (Shum & Luckin, 2019), and simulation-based learning (Dai et al., 2021; Winkler-Schwartz et al., 2019). Among these educational applications using AI, simulations have been widely recommended and used to provide active and adaptive learning experience (Chen et al., 2020; Johnson & Lester, 2016). As such, in this review, we focus on exploring and synthesizing the affordances of AI in addressing individual needs (Chen et al., 2020; Woolf et al., 2013) and providing active learning experiences in simulation-based learning.

Simulation-based learning environments, such as intelligent systems, virtual reality, or medical simulations, refer to "interactive digital learning environments that imitate a real-life process or situation…allow learners to test their hypotheses of the effects of input variables on the intended outcomes" (Merchant et al., 2014, p. 30). By using AI techniques (such as artificial neural networks (ANNs), natural language processing (NLP), machine learning, and deep learning), simulation-based learning can facilitate active learning and decision making with human-like interactions. However, the literature remains scarce concerning the trends in the applications of AI (Chen et al., 2020) in simulation-based learning——specifically, the associated aspects of active learning experiences and strategies (Hwang et al., 2020; Zhang & Aslan, 2021). As a result, the purpose of this systematic review is to conduct literature mapping and analyze thematic trends of *educational applications of AI in simulation-based learning* (or “AI in simulation-based learning” hereafter). We also focus on the associated learning principles and theories underlying the trends in AIEd research. Particularly, the research question guiding this systematic exploration is: *What are the trends of educational applications using AI in simulation-based learning?*

1.1. Defining AI

“Can machines think,” the pioneering question proposed by Turing (1950, p. 433) illuminates human curiosity in conducting research in machine intelligence. AI, a term first used in 1956, is based on the conjecture that “every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it” (McCarthy et al., 2006, p. 12). A more recent definition of AI can be stated as “computers which perform cognitive tasks usually associated with human minds, particularly learning and problem-solving” (Loder & Nicholas, 2018, p.11). Accordingly, to solve contemporary educational problems by simulating human minds, modern AIEd techniques include three facets (Chen et al., 2020)——deep learning (e.g., deep neural networks, recurrent neural networks, long short-term memory), machine learning (e.g., reinforcement learning, decision trees, Bayesian networks), and AI (e.g., fuzzy logic, rule-based systems, agent system). Aside from deep learning and AI, as Chen et al. (2020) elucidated, it is widely acknowledged that intelligent agents are trained by machine learning techniques. In alignment with the aforementioned study focus, intelligent agents play an important role in addressing individual needs in simulation-based learning as a form of AI (Johnson & Lester, 2016). Therefore, we included "artificial intelligence" and "machine learning" as well as related terms of "intelligent agent" in our literature search to reflect the study focus (see Table 1).

1.2. Defining simulation-based learning

Simulation is defined broadly in this review as "a practical experience that produces a convincing re-creation of a real-life event or set of conditions" (Alinier, 2007, p. 248); learners have to be actively involved in communicating and interacting in the immersive environment to solve a problem (Miller, 1984). Taken together, we define simulation-based learning as—an active and immersive experience in a virtual learning environment that re-create a realistic and authentic real-life event or a set of learning situations to solve a problem. Simulation-based learning can afford a cost-effective authentic learning experience that facilitates deeper learning, inquiry, and problem-solving. For example, medical science is a domain where simulation-based learning has been applied intensively, especially for surgical training (Alinier, 2007; Castillo-Segura et al., 2021; Winkler-Schwartz et al., 2019). The learner experience of simulations based on presence and realism can differ. Virtual reality, augmented reality, mixed reality, virtual lab, serious games, or computer-based simulation systems are all under the umbrella term of simulation-based learning environments (Gandedkar et al., 2021; Merchant et al., 2014) in this systematic review.

2. Method

To develop informative insights in this review, we adopted a systematic mapping review for literature selection, review, analysis, and report (Petersen et al., 2015). To supplement nuanced details in the literature, we also conducted a thematic synthesis. According to Petersen et al. (2015), a systematic mapping review is useful to understand the structure and research agenda of a topic area, whereas thematic synthesis integrates the strengths of the evidence. This review has a strong focus on educational applications of AI and how different aspects of learning (i.e., productive inquiry and learning, affective states associated with learning, and learning assessments) are potentially promoted or influenced in the technologies designed (i.e., AI and simulation-based learning).

We searched several academic databases, and the initial search resulted in a total of 1,695 records: APA PsycInfo (n=481), Web of Science (n=394), Science Direct (n=306), ERIC (n=65), ACM Digital Library (n=434), and IEEE Xplore (n= 15). Google scholar and the reference list of the retrieved papers were also used for additional literature search. The search terms were listed in Table 1 with the associated topics and relevant literature.

Table 1.

Topics, relevant literature, and mapping search terms used in this review

Topics and relevant literature	**Mapping of search terms used in this review**
AI (e.g., Chen et al., 2020; Guo & Goh, 2015; Rickel & Johnson, 1998)	(“artificial intelligence” OR “machine learning” OR “conversational agents” OR “pedagogical agents” OR “chatbots” OR “autonomous agents” OR “embodied intelligent agent” OR “disembodied intelligent agent”) AND
Simulations (e.g., Merchant et al., 2014)	(“virtual reality” OR “intelligent tutoring systems” OR “game-based learning” OR “learning game” OR “educational game” OR “serious game” OR “simulation” OR “simulated learning” OR “simulation-based learning” OR “immersive learning” OR “augmented reality” OR “mixed reality”) AND
Education/learning	(“education” OR “learning”)

To analyze broader trends, the studies reviewed comprehensively included empirical, conceptual, and design-oriented technical papers published in peer-reviewed venues. We excluded unpublished papers and sought to identify peer-reviewed papers that met the inclusion criteria to ensure the study quality (Wohlin et al., 2013). Table 2 summarized the inclusion and exclusion criteria. After title and abstract screening and inclusion and exclusion criteria applied, 59 articles were eventually included, reviewed, and analyzed at the full text level with a focus on the trends of educational applications using AI in simulation-based learning. Figure 1 presented the systematic process for literature selection.

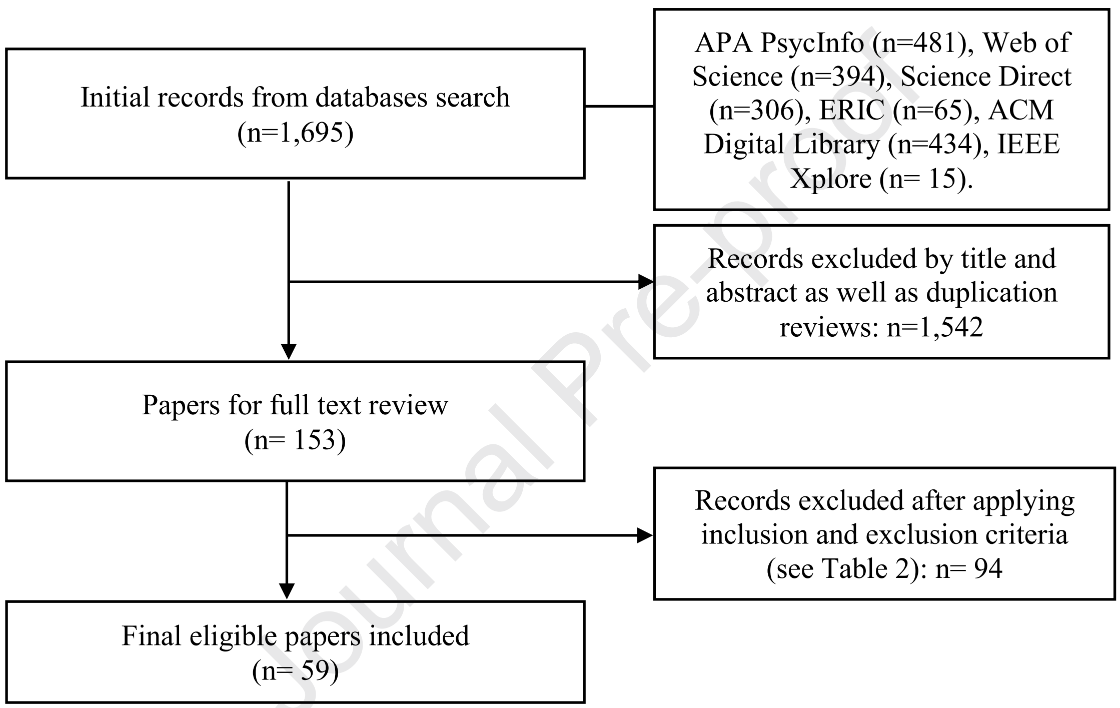


Figure 1.
Systematic process for literature selection in this review

Table 2.
Inclusion and exclusion criteria

Criteria type	Description	Notes
Inclusion	1. Peer-reviewed journal articles, book chapters, and conference proceedings 2. Topic relevance: AI 3. Topic relevance: simulations 4. Topic relevance: education/learning 5. In English	To be included, the articles need to meet all the inclusion criteria.
Exclusion	1. Not peer-reviewed journal articles, book chapters, and conference proceedings	If one of the exclusion criteria

	2. Topics irrelevance: not AI 3. Topics irrelevance: not simulations 4. Topics irrelevance: studying stakeholders perceptions toward AI 5. Topics irrelevance: studying how to teach AI 6. Topics irrelevance: not related to education/learning 7. Not in English	is met, the article would be excluded.

Among all the studies of AI in simulation-based learning we included via a systematic literature synthesis, three thematic trends were identified. After the initial open coding, we constantly refined and congregated the codes to better capture the emerging thematic trends. AI was used in simulation-based learning as manifested in the three broad themes identified: *Virtual agents, Affects*, and *Assessments*. A coding matrix (as in Table A1 in the Appendix) was then used to organize the data and map the results (Wohlin et al., 2013). In addition to the three themes, we also coded *Methods, AI technologies, Simulation [Learning principles or theories], Learning outcomes* (mostly applicable for empirical studies and meta-analyses), and *Research issues and contributions* in the coding matrix using Excel spreadsheets. These codes were central to our research question. We defined essential codes and their subcodes (i.e., *Methods*, *AI technologies*, *Simulation*, and *Learning principles or theories*) in Table 3. A trained coder read the included studies in entirety and identified the code-relevant information explicitly mentioned by the study authors or fairly interpretable by the coder. We established credibility by repeatedly comparing the codes and the studies as well as by keeping analytical memos (Miles et al., 2020).

Table 3.

Definition and description of the codes and subcodes

Codes ・Subcodes	**Descriptions**
Methods	Research design for the investigation. Unique to AI studies, we also identified machine learning validation or user experience studies.

▪ Conceptual or architecture design	Presenting new approaches, ideas, models/architectures for educational application (non-empirical); e.g., Johnson & Lester (2016).
▪ Experiments	Evaluating and examining the effects of the designed intervention with scientific procedures; e.g., Kumar & Rosé (2011).
▪ Systematic literature review or meta-analysis	Reviewing literature (systematic literature reviews) or synthesizing effect sizes (meta-analyses).
▪ Usability or evaluation study	Using software evaluation methods to study aspects such as usability, efficiency or user's experience; e.g., Flores et al. (2013).
▪ Machine learning validation	Adopting machine learning methods to evaluate the applications in simulations (Wang & Zheng, 2013; Myllyaho et al., 2021); e.g., Westera et al. (2018) used machine learning to validate automatic scoring prediction.
▪ Qualitative (case study or interpretive study)	Studying the phenomenon in the natural settings to develop useful insights; e.g., Dalinger et al. (2020).
▪ Mixed-methods design	Employing both qualitative (e.g., observations and/or interviews) and quantitative (e.g., experiments and/or questionnaires) techniques; e.g., Lorenzo et al. (2013).
AI technologies	Following our definition of AI in Section 1.1., we identified the AI technologies categorized below in the included studies.
▪ Machine learning algorithms or ensemble methods	Machine learning, deep learning, or AI algorithms used; sample keywords include decision trees, Bayesian networks, fuzzy logic, Convolution Neural Networks (CNNs) (Chen et al., 2020), or adaptive algorithms (Aleven et al., 2017). Ensemble methods suggest that the study combined more than one algorithm (Opitz et al., 1999).
▪ Module-based modelling	Featuring popular AI modelling technologies in simulation-based learning including student modelling, agent modelling, knowledge modelling, teacher modelling, or Open Learner Modelling (OLM) (e.g., Gandedkar et al., 2021; Kumar & Rosé, 2011).
▪ Hybrid mechanisms	A combination of knowledge modelling and machine learning (e.g., Aleven et al., 2017).
▪ NLP or text-to-speech (TTS) conversion technology	Although it can be a subset of machine learning (Jordan & Mitchell, 2015), natural language is an important aspect of simulation-based learning interactions thus we coded as an individual category, including studies explicitly used keywords of NLP or TTS.
▪ Rule-based AI	One of the predetermined AI technologies with lower sensitivity and responsiveness. Rule-based AI technologies "use algorithms either *handcraft* or *already existing ones* as a decision maker to identify both knowledge and response" (Maroengsit et al., 2019, p. 112).
▪ AI markup language	The XML (also called *Extensible Markup Language*) developed for supervised, stimulus-response chatbots (Wallace, 2009).
▪ Scripted AI	Referring to "lists of rules that are executed *sequentially*" (Spronck et al., 2006, p. 218), which is similar but less dynamic than rule-based AI.
▪ Affective computing	Computational approaches that associate, detect, express, or purposefully influence emotions (Picard, 1997).

▪ Multimodal technologies	Computational systems that detect, interact, assess or evaluate human perceptions, activities, and interactions (Stiefelhagen et al., 2007).
▪ AI in general without technical details	Articles of simulation-based learning that only mentioned general keywords such as “AI,” “machine learning,” or “deep learning” but did not provide details or explanations.
Simulation	Following the definition in Section 1.2., we found the following simulation categories in the included studies.
▪ Intelligent tutoring systems (ITSs)	Digital learning environments that incorporate computational models to tutor learners and track their psychological states (Graesser et al., 2012).
▪ Virtual/mixed reality	Both virtual reality and mixed reality use computer graphics systems. Virtual reality combines different display and interface mechanisms to constitute an immersive and interactive 3D environment; whereas mixed reality incorporates real/authentic 3D scene or real world elements into the virtual environment (Pan et al., 2006).
▪ Simulation games	An artificial environment that situates and immerses learners in decision-making scenarios, learners learn from the consequences of the decisions they made (Sitzmann, 2011).
▪ Medical simulation	Replicating medical and surgical procedures or demonstrating clinical features with computerized virtual and life-like devices for assessment, diagnosis, or clinical practices (Kunkler, 2006).
▪ Smart edutainment	Conceptualized from smart education/learning in which learners construct knowledge and develop competences with adaptive and personalized edutainment. Only Guran et al. (2020) in the included studies adopted this term.
Learning principles or theories	We define this category as systems of ideas, concepts, and frameworks for understanding and explaining how humans learn.
▪ Cognitive theory of multimedia learning	Incorporating evidence-based principles for the designs of computer graphics, videos, and instructional texts that help humans learn better (Lawson et al., 2021; Mayer, 2020).
▪ Motivation theories on learning	Motivation is “the processes that instigate and sustain goal-directed activities” (Schunk & DiBenedetto, 2020, p. 1). The included studies may use *motivation* as a keyword or design simulation-based learning with the ARCS (Attention, Relevance, Confidence, and Satisfaction) model of motivation (Keller, 2010).
▪ Social learning and social constructivism	Concepts of co-constructing knowledge, providing feedback, scaffolding, or zone of proximal development (Vygotsky, 1978) that were explicitly mentioned in the included studies.
▪ Self-regulated learning	Learners are active participants that take responsibilities to systematically plan, guide, and master their own learning (Zimmerman, 1990).
▪ Situated cognition	Assuming a dynamic and unique relations between learners and the learning environments in which in-situ, contextualized, and life-like learning occurs to cultivate the learning process (Choi & Hannafin, 1995). Concepts and keywords include apprenticeship, coaching, real-situations, situated practice, transfer, or contextualized learning.

・Social cognitive theory	Based on Bandura's (1986) reciprocal interactions include three processes: personal (sample keyword: self-efficacy), behavioral (sample keywords: effort, persistence), and environmental (sample keywords: observing or modelling others, vicarious learning) (Schunk & DiBenedetto, 2020).
・Collaborative learning	Learners can collaborate with AI-powered peers or with human peers facilitated by AI for the learning experience (e.g., Biswas et al., 2005; Terzidou et al., 2016).

3. Results

We presented the coding results of 59 studies in Table A1 in the Appendix. For the results, we first reported the mapping study, and then presented the synthesized thematic trends. For the literature mapping in Section 3.1., we summarized (1) the year-of-study distribution, (2) methods, (3) mapping of AI technologies, (3) mapping of simulation, (4) mapping of study trends, and (5) mapping of the underpinning learning principles and theories. Via a systematic literature synthesis, we identified three thematic trends in Section 3.2.: (1) AI built in virtual agents for simulation-based learning, (2) AI infused in simulation-based learning with affective computing, and (3) AI leveraged in simulation-based learning for assessments.

3.1. Results of the mapping study

3.1.1. The year-of-study trend

Figure 2 showed the year-of-study trend. The year range of the included studies was 1998 to 2021. With the most included studies (n = 10) published in 2021, followed by 2020 (n = 8), demonstrating an upward trend in research publication since 2019.

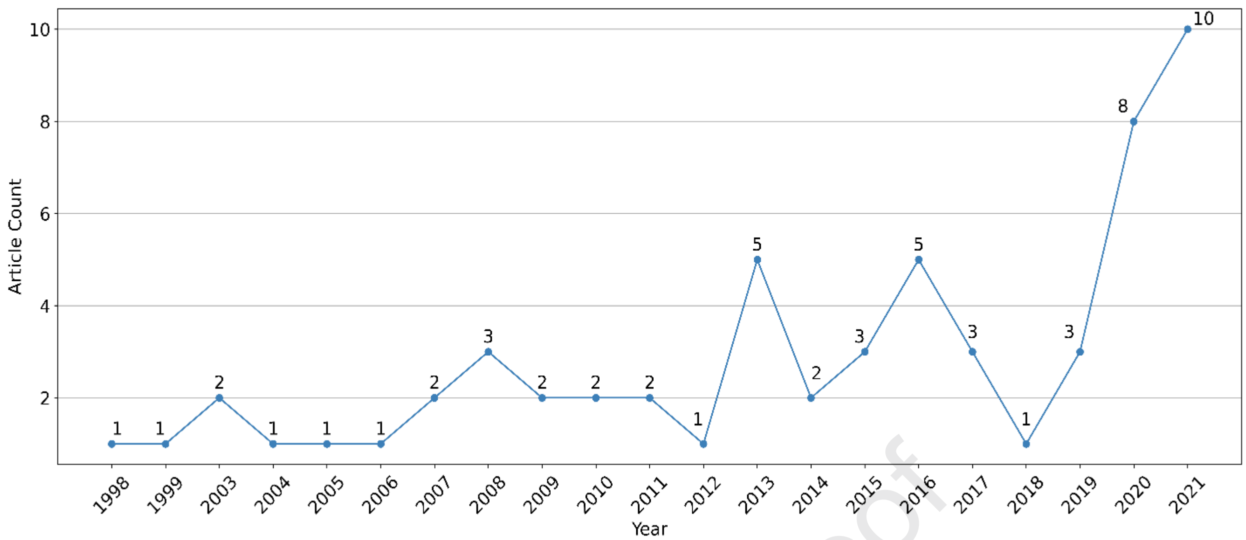


Figure 2.

The year-of-study trend during the period of 1998-2021

3.1.2. Methods

We charted the methods used in the included studies (see Figure 3). Due to the nature that simulation-based learning with AI applications is in its nascent stage, conceptual or architecture design papers were the most frequent in the included studies (n = 29) followed by experiments (n = 20), systematic literature review or meta-analysis (n = 10), usability or evaluation study (n = 9). Further, machine learning validation or qualitative studies accounted equally (n = 5). There were 2 studies that adopted mixed-methods design.

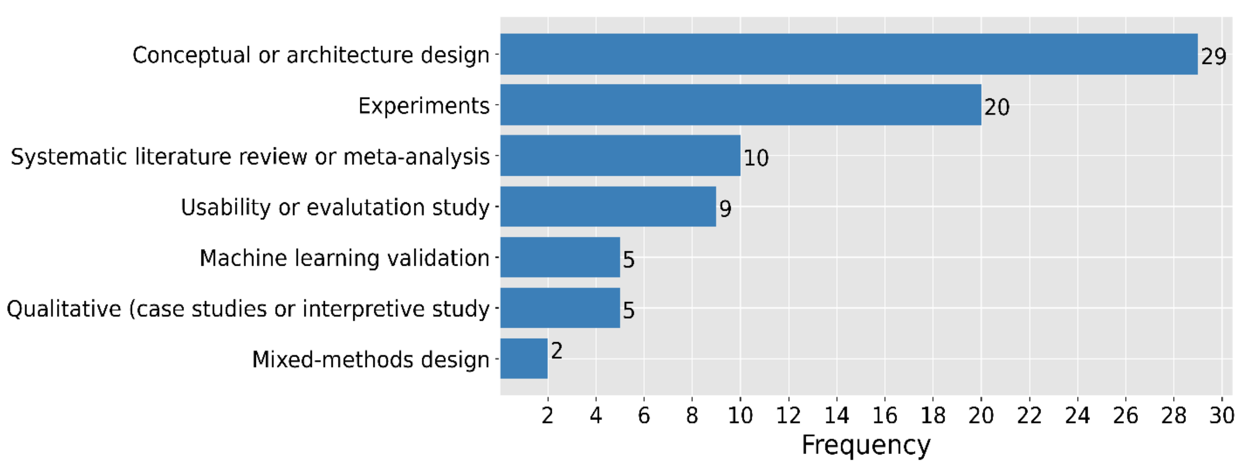


Figure 3.

Distribution by methods

3.1.3. AI technologies

There were seven categories for AI technologies used in the literature (see Figure 4). *Machine learning algorithms or ensemble methods* dominated the included studies (n = 29). *Module-based modelling* (n = 25) was also frequently featured among simulation-based learning to model modularized knowledge and intelligence. One study (Aleven et al., 2017) used *hybrid mechanisms* for adaptive systems to provide personalized learning experience by adapting to multiple psychological states. Moreover, NLP or TTS conversion technology (n = 15) was useful for making authentic and naturalistic interactions in simulation-based learning. Other types of techniques, representing more pre-determined interactions (compared to using machine-learning-driven interactions), were rule-based AI (n = 3), AI markup language (n = 1), and scripted AI (n = 5). *Affective computing* (n = 7) was used to model and express emotions, sentiments, and affective states from the computers; or to assess emotions, sentiments, and affective states from the learners. Technologies of sensorimotor or detector-enabled multimodal interactions, including body movements, in-simulation learning analytics or surgical training, were also identified (n = 5). There were a few studies that only mentioned AI in general without technical details (n = 4).

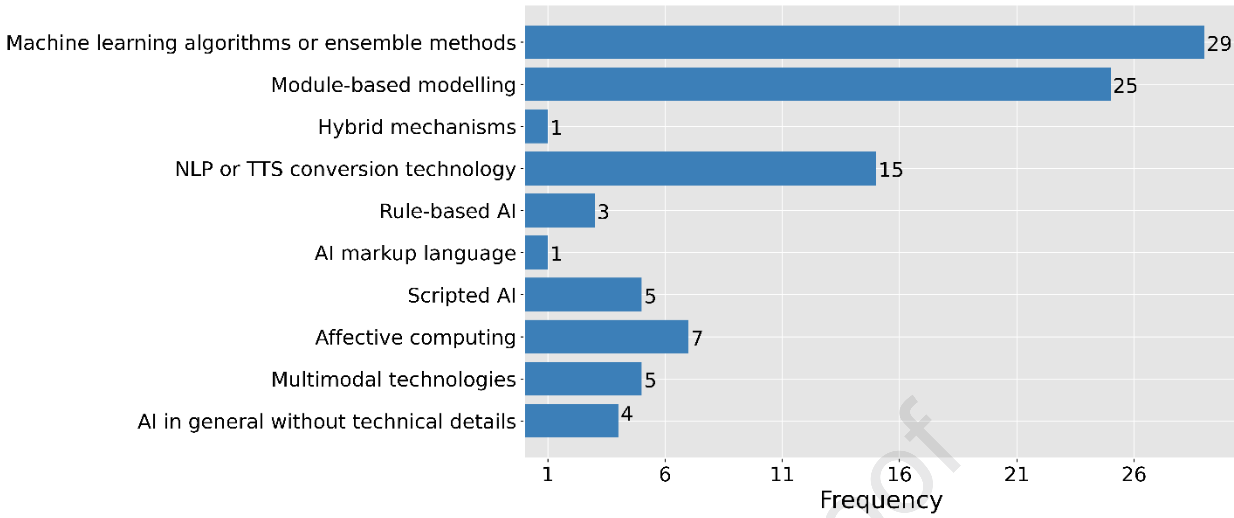


Figure 4.

Distribution by AI technologies

3.1.4. Simulation

The dominating simulation-based learning environments used in the coded studies were ITSs (n = 40, 66%). Different studies named ITSs differently. In Figure 5, ITSs category included simulations such as web-based simulation, interactive simulation microworld, or virtual laboratory. Further, virtual reality or mixed reality (n = 9, 15%) and simulation games (n= 9, 15%) were equal in the mapping of simulation. Smart edutainment and medical simulation accounted for 1.6% (n = 1) and 3.3% (n = 2) respectively.

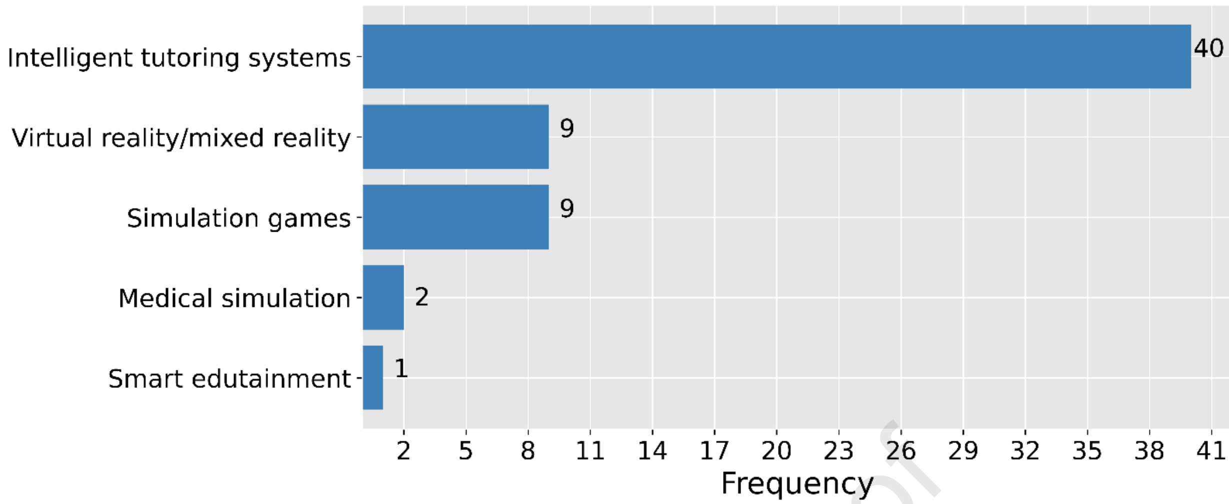


Figure 5.

Distribution by Simulation

3.1.5. Study trends

We observed study trends in three fundamental areas inferred from literature synthesis: Virtual agents, Affects, and Assessments. Assessments (n = 16) and Virtual agents (n = 14) were the two most studied areas using AI in simulation-based learning. Figure 6 displayed the three trends and the interlinked areas (i.e., Assessment/Virtual agents, Affects/Assessments, Virtual agents/Affects). Interestingly, the studies with affective computing (i.e., Affects) in this review were interlinked with Virtual agents and Assessments (see Section 3.2.2. for more descriptions), there was no study focused solely on Affects (n = 0).

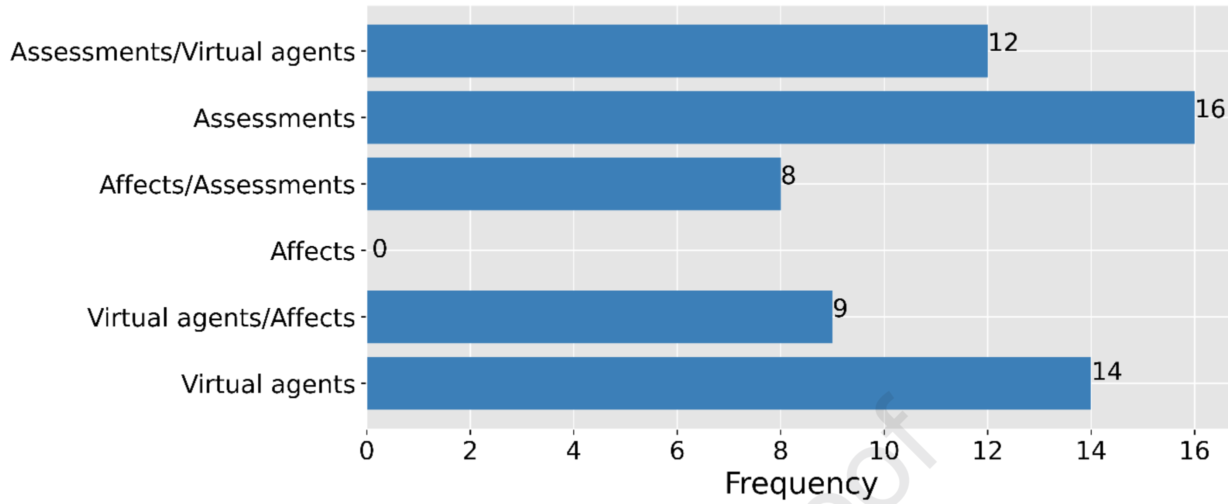


Figure 6.

Distribution by Study trends

3.1.6. Study trends and learning principles and theories

Seven contemporary learning principles and theories were identified and mapped in the coded studies (see Figure 7 for more details). The most frequently mentioned ones were Self-regulated learning as the foundation of Assessments (n = 8) and Social learning and social constructivism that supported the design of Virtual agents (n = 8). Virtual agents were also used for Motivation purposes in the included studies (n = 7), sharing the same frequency with Virtual agents/Affects and Assessments/Virtual agents. As the major theory for motivational processes, Social cognitive theory (Schunk & DiBenedetto, 2020) was the common guiding theory of the studies on Virtual agents/Affects (n = 3), Affects/Assessments (n = 3), and Assessments/Virtual agents (n = 3). The learning principles and theories used were totaled as follows: Motivation theories on learning (n = 29), Social learning and social constructivism (n = 23), Self-regulated learning (n = 21), Situated cognition (n = 12), Social cognitive theory (n = 8), Collaborative learning (n = 4), and Cognitive theory of multimedia learning (n = 3).

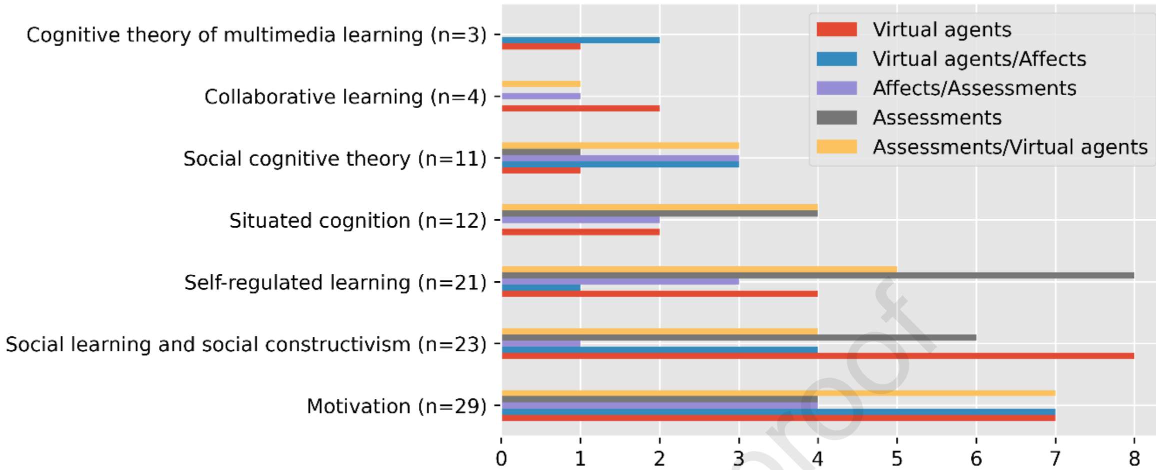


Figure 7.

Distribution by Study trends and learning principles and theories

3.2 Results of thematic literature review

3.2.1. Trend 1: AI built in virtual agents for simulation-based learning

The first trend unveiled a prominent feature of using AI-powered virtual agents in simulation-based learning for active learning. AI agents accounted for 59% (n = 35) in Figure 6 (including Assessments/Virtual agents, Virtual agents/Affects, and Virtual agents). Using AI-powered virtual agents could address one of the grand challenges in education——providing personalized learning (Woolf et al., 2013).

Self-regulated learning (Zimmerman, 1990) is one of the most prominent aspects of learning that AI-powered virtual agents support (Graesser & McNamara, 2010). Productive learning can occur when a learner reflects on and plans for in-simulation learning acts (Perez et al., 2017). A successful self-regulated learner can systematically apply learning strategies to set goals, identify their own gaps in knowledge and skills, ask quality questions, select, interpret, and (re)organize information (Graesser & McNamara, 2010; Zimmerman, 1990).

There are different approaches to supporting self-regulated learning. OLM is one of the approaches that has been promoted (Gandedkar et al., 2021; Nussbaumer et al., 2015). OLM can employ machine learning techniques, Bayesian networks, fuzzy logic, or item response theory (Hooshyar et al., 2020) to make learners' real time performance "open" to themselves so that the learners can plan, monitor, and reflect on their own learning for self-guided decision making (Nussbaumer et al., 2015; Graesser & McNamara, 2010). Graesser and McNamara (2010) suggested that the OLM in self-regulated learning can be achieved through natural language interactions between the learners and virtual agents. Natural discourse provides powerful adaptivity to guide student learning.

Grounded in situated cognition, AI-powered virtual agents that support productive inquiry are sometimes referred to as *guidebots* (Johnson, 2003a). These *guidebots* support multiple learner-agent interactions, including hints and feedback as cognitive assistance, positive messages as motivational prompting, and gestures as emotional and enthusiastic companions (Johnson, 2003a; 2003b). Notably, the cognitive assistance of the *guidebots* was designed to guide inquiry and thinking rather than providing answers (Hwang et al., 2020). These efforts require the design and development of student knowledge modelling and evidence-based reasoning in AI-powered virtual agents (Johnson, 2003b). The purpose of these mechanisms is for the AI-powered virtual agents to guide learners to be aware of, reflect on, and correct their misconceptions during the learning processes. For example, the AI-powered virtual agent of *Adele* in Johnson (2003b) was equipped with Bayesian networks to automatically update the dialogue based on learners' performance, and guide learners' reasoning processes for knowledge construction. Similarly, de Freitas and Neumann (2009) articulated *exploratory learning* as a theoretical underpinning for immersive simulation-based learning. AI-based systems should support students' problem-solving skills development via the cycle of experience, explore, reflect, abstract, and test (de Freitas & Neumann, 2009).

Although productive inquiry and learning with AI-powered virtual agents can possibly facilitate knowledge construction when learning in contexts, another consideration is that it may not be equally beneficial for all learners. Some research maintains that virtual agents do not work well when learners have high level of prior knowledge (Graesser et al., 2008) since the role of AI agents is mainly to facilitate "expectation- and misconception-tailored dialogue" (Graesser et al., 2008, p. 302). Students in their zone of proximal development may yield the most productive learning, whereas students with either higher or lower prior knowledge would not enjoy or engage in learning with AI-powered virtual agents.

Closely related to self-regulated learning discussed above, *motivation* (Zimmerman, 1990) is a central aspect that AI-powered virtual agents strive to support (Gulz & Haake, 2006; Kim & Baylor, 2016; Lippert et al., 2020; Terzidou et al., 2016). Lippert et al. (2020) revealed that using trialogues with different AI-powered virtual agents (as a teacher and a peer) has the potential to support motivation. For example, human learners can learn from AI peer's mistakes without sacrificing their self-efficacy or motivation to learn. In the experiment comparing the use of virtual agents (called *Emma*) and without virtual agents (control condition), Van der Meij et al. (2015) found that girls benefitted from *Emma*, an agent designed based on the ARCS model of motivation (Keller, 2010). Indeed, as Gulz and Haake (2006) pointed out, questions regarding motivation in AI-powered virtual agents should be "for which people, in which conditions and for what kinds of domains that a certain effect can be shown" (p. 325), rather than the whether-or-not questions.

Learning by teaching with AI-powered virtual agents is another technique of AI-powered simulation-based learning. As a social/collaborative learning entity, AI-powered virtual agents are *teachable* (Biswas et al., 2005; Chin et al., 2010; Matsuda et al., 2020). Importantly, compared to the aforementioned *guidebots* that benefit learners with lower prior knowledge, teachable agents arguably benefit students with higher prior knowledge more because interacting with a teachable agent requires complex problem-solving skills and deeper domain knowledge (Johnson & Lester, 2016; Nye et al., 2014). However, scaffolding can be added to the learning system integrating teachable agents. Specifically, metacognitive scaffolding that helps learners monitor or identify their teaching strategies was found to be equally effective among students with diverse prior knowledge (Matsuda et al., 2020). Prior studies reported that teachable agents equipped with self-regulated learning supports would better prepare students for future learning (Biswas et al., 2005; Chin et al., 2010).

Finally, Veletsianos and Russell (2013) synthesized multiple roles of virtual agents, including conversational partner, instructional expert, information broker, expert system, and playmate. However, they also argued that poor syntax and lack of context in the usage of virtual agents can result in confusion. Recently, AI-powered conversational agents have improved in its communicative ability thanks to large transformer-based language models trained with big data (Brown et al., 2020). Open-source agents with long-term memory can provide more authentic context-rich interactions (Xu et al., 2021). Nonetheless, these cutting-edge agents are not trained with educational context-specific data; using data in educational context to train or extend these agents, created with large transformer-based language models, is in need of research (Dai et al., 2021). Moreover, the underlying learning principles and theories that guide the use of AI-powered virtual agents should be further explicated.

3.2.2. Trend 2: AI infused in simulation-based learning with affective computing

In the literature, the sources of affects were considered bi-directional, either from the learners or from AI technologies. Thus, in the mapped categories, there was no study focusing on affects *alone* (see Figure 6). Sensing, detecting, and classifying affective states of the learners and AI-powered virtual agents require multimodal data capturing and mining, such as facial detection, semantic analysis or emotion recognition. Guiding theories for integrating affective computing in simulation-based learning are motivation theories on learning and social cognitive theory (Schunk & DiBenedetto, 2020) (see Figure 7). In this section, we first discussed how AI can be used to detect and classify learners' affective states during learning, we then presented how AI-powered virtual agents can be used in simulation-based learning to convey emotions/affects to the learner.

Given the interdependent relations between affects and learning, there is increasing research on designing and developing mechanisms in AI simulation-based learning to track learners' affects. Prior research has employed automatic emotion recognition via a deep learning approach to support young children in simulation-based learning (Guran et al., 2020). Guran et al. (2020) applied supervised machine learning with CNNs to classify basic emotions (i.e., neutral, sad, and happy). A prominent challenge of such an approach was the lack of a big sample of young children's images due to protection procedures with minors. The authors opted to use a mixed image dataset of both adults and children, and the transferred results with the benchmark dataset can bring the best accuracy rate (81%) for children's emotion recognition. Similarly, earlier work (Graesser et al., 2008) suggested that learners' emotions classified by computer can reach 65% to 70% accuracy. The accuracy would be higher if multimodal data such as facial detections, body posture, and discourse intonation were used (Graesser et al., 2008).

Classifying learners' emotions or affective states (D'Mello & Graesser, 2012) in simulation-based learning is important because cognition, learning, and emotions are interlinked. Affective-sensitive AI systems in simulation-based learning can track students' affective states (D'Mello & Graesser, 2012). D'Mello and Graesser (2012) introduced *Affective AutoTutor* using state-of-the-art sensing technologies and machine learning techniques to address and regulate students' boredom, confusion, and frustration by capturing and mining their conversational cues, body language, and facial features. Their experiments with computer literacy undergraduate students reported that students with higher prior knowledge did not benefit from emotional support AI-powered agents; for low prior knowledge students, AI-powered affective agents were more effective later in a lesson than at its beginning.

Another aspect of affective computing in AI-powered learning simulations is using emotion-expressive virtual agents (Guo & Goh, 2015; Rickel & Johnson, 1998). Embodied agents in simulation-based learning can cue learners via virtual motor actions, gaze and gesture to direct students' attention and convey verbal and non-verbal communications (Rickel & Johnson, 1998). Guo and Goh's (2015) meta-analysis found that AI-powered affective agents had a significant and moderate effect on motivation to learn, and yielded a small effect on knowledge retention and knowledge transfer. Positivity principle was also applied to embodied agents in simulation-based learning (Lawson et. al., 2021). Lawson et al. (2021) revealed that learners were able to attend to virtual agents' emotional expressions (positive or negative), and self-reported that virtual agents with positive emotions promoted learning. Although learners noticed virtual agents' positive emotions more and better enjoyed the instruction than learning from virtual agents with negative emotions, a delayed test indicated that virtual agents with positive emotions did not significantly differ from the negative-emotion counterpart (Lawson et al., 2021). Future research is needed to further study the delayed effects. Voice is another cue for emotions. In their experiments, Craig and Schroeder (2017) compared virtual agents with voices using classic speech engine, contemporary speech engine, and human voice. They found a voice effect for science learning (i.e., the formation of lightning). Participants in the contemporary TTS engine group outperformed classic and human voice groups in the transfer test.

3.2.3. Trend 3: AI leveraged in simulation-based learning for assessments

Assessment is regarded as a key component in AI-integrated simulation-based learning (Castillo-Segura et al., 2021; Dalinger et al., 2020). For example, Dalinger et al. (2020) used a mixed reality simulation (i.e., Mursion) integrated with AI- and human-controlled virtual students for preservice teacher training. They argued that assessment in such a simulation facilitated participants to reflect on their gap in teaching performance. Using a mixed reality simulation, Ke et al. (2016) incorporated a blend of human peers and AI-scripted virtual students in the assessment of teaching efficacy. Real-time feedback and multimodal interactions facilitated the teaching performance assessment (Dalinger et al., 2020; Ke et al,. 2016). Growing attention has been given to in-situ simulation-based learning analytics (Heradio et al., 2016). Innovative ideas include recording in-situ learners' behaviors as well as utilizing an AI learning analytics dashboard. Although in-situ simulation-based learning analytics is still limited in its applications, Heradio et al. (2016) expected more research in this area in the near future.

Current practice of AI assessments for supporting student learning involves the use of data-driven prediction. Modern machine learning algorithms used to support the assessment for learning are Support Vector Machine (SVM), Neural Networks (Castillo-Segura et al., 2021;

Moon et al., 2020; Yang et al., 2019), and latent semantic analysis combined with regular expressions (Graesser et al., 2017). Castillo-Segura et al.'s (2021) systematic literature review on surgical technical skills assessment revealed four aspects to be considered——(1) sensors, (2) multimodal data collection (i.e., body movements, eyes tracking, and/or the combinations) including baseline indicators, (3) machine learning algorithms (i.e., SVM and Neural Networks), and (4) feedback mechanism (i.e., real time and delayed feedback). Likewise, Hidden Markov Models (HMMs) were used for assessing, classifying, and providing feedback on dental skills in simulation (Rhienmora et al., 2011). In-simulation data were mined to afford the automatic tracking of users' actions. The assessment afforded by HMMs was then used to generate tutoring feedback for the dental operations. Yang et al. (2019) built ANNs computational models to assess learners' cognitive abilities in English language learning. ANNs simulate human brains by interconnecting neurons to process information simultaneously. The architecture of ANNs consists of input, hidden, and output layers. The three-layer ANNs model results can be used to predict performance or to map features for the prediction. Yang et al.'s (2019) ANNs model results were applied in a user-friendly online platform to predict learners' cognitive performance and assess learners' *verb* and *past tense* performance. Yang (2009) suggested that the generated reports helped teachers and instructional designers identify learners' learning progress to adapt teaching.

Further, to supplement Castillo-Segura et al.'s (2021) guideline for multimodal assessment, we also presented a learning-task-relevant protocol. Lorenzo et al. (2013) used the supportive tasks protocol (called TEVISA) as an integrative approach to support learning, training, and assessment in immersive virtual reality. The protocol consists of executive function and social competency training for learners with Asperger syndrome. In addition to employing avatars with AI to provide adaptive intervention when learners were not engaged in the learning task, the authors also endorsed multimodal assessment by tracking body coordination, voice, eyes, attention, and empathy controls with different technologies such as the Dynamic Time Warping Kinect SDK (DTW) Gesture tool and other facial and voice recognition technologies.

4. Discussion

4.1. Mapping study

Via the literature mapping, we identified a growing trend on the research of simulation-based learning in AIEd since 2019. Capturing the nature that modern AI in simulation-based learning is still growing, most studies included were conceptual or architecture design articles. Unique to AI in simulation-based learning, we also identified studies that used usability/evaluation methods or machine learning validation methods. These studies were useful for the development of AI in simulation-based learning. However, to realize the effectiveness and efficacy of AI in simulation-based learning, more experimental and quasi-experimental studies should be conducted. Importantly, the simulation-based learning environments using AI that are researched in the experimental studies should involve systematic and iterative design and development endeavors driven by usability, evaluation, or machine learning validation studies.

We found that most of the included studies used ensemble methods in machine learning or module-based and hybrid modelling. This finding is in partial agreement with prior research in that learner models for explanatory and actionable insights are critical rather than focusing on machine learning accuracy (Rosé et al., 2019). To elaborate, module-based and hybrid modelling takes advantage of models such as student modelling, knowledge modelling, or OLM for the purpose of enhancing human learning by explaining the learning process and adapting to an individual’s needs. These modelling techniques coupled with machine learning are unique characteristics of AI applied in simulation-based learning, in comparison to AI used in other educational settings such as online learning because these models shape the interactions in simulation-based learning. In terms of simulation, ITSs were widely used in the included studies. Possible reasons are that ITSs are web-based and thus easier to be implemented in research settings or classroom instructions compared to other types of simulation. Research on virtual/mixed reality and simulation games is lacking in the present review, and is an important area for future investigation because mixed/virtual reality and simulation games can be more immersive; more notably, they can provide more in-situ practices for knowledge and skills application of which modern AI can be used with substantial potential (Chen et al., 2020; de Freitas & Neumann, 2009).

We identified three major study trends (i.e., virtual agents, affective computing, and assessments) and their interlinked relations. The interlinked relations point to the importance of collaborations between multiple disciplines such as educational technology, computer science, cognitive science, and psychology for the design and development of effective simulation-based learning applying AI. Finally, it is critical to be cognizant that the focus of AI used in educational applications should undoubtedly be on enhancing human learning based on evidence-based learning principles and theories. In this review of AI in simulation-based learning, the category of *motivation theories on learning* is the most frequently used, followed by *social learning and social constructivism* and *self-regulated learning*. We encourage AIEd researchers and practitioners to use the findings of this review for designing AI-powered simulation-based learning with the underpinning learning principles and theories that fit into their specific learning contexts (see Figure 7).

4.2. Thematic synthesis

Virtual agents remain one of the crucial aspects of AI in simulation-based learning. Self-regulated learning, motivation, situated cognition, and social learning are the background learning principles and theories in this entity. There is a general acknowledgement that the functions of virtual agents should not be the knowledge provider but rather a guide or companion so that the learning experience is self-initiated and constructive. That is, learners are expected to be involved in reasoning and problem-solving processes through productive interactions with the agents (de Freitas & Neumann, 2009; Graesser & McNamara, 2010; Johnson, 2003b). In this highly interactive and demanding learning context, prior knowledge has been a debating topic in the research of virtual agents (Graesser et al., 2008). More recent studies suggest that adding scaffoldings may help to satisfy learners with different levels of prior knowledge (Matsuda et al., 2020). Nonetheless, the effects of various scaffoldings used in simulation-based learning with AI-powered virtual agents are still in need of more research, especially when AI responses are stochastic with predictive machine learning models. Another concern is the fidelity and authenticity of the human-AI conversations. This concern poses challenges for creating authentic adaptive learning. The review results suggested the needs for innovative development in modern AI to drive adaptive or personalized simulation-based learning (Chen et al., 2020; Woolf et al., 2013).

Affective states play a key role in simulation-based learning using AI. Learners' affective states and how learners being influenced by the affects expressed by AI can both impact learning trajectories as it potentially regulates students' learning behaviors. Research in this review on affective computing for simulation-based learning connects closely with virtual agents and assessments. It is warranted for future research to investigate affective states *coupled with* virtual agents and assessments for meaningful learning experience. Affective-sensitive AI systems are critical in simulation-based learning because learners are prone to frustration or confusion in such self-initiated and exploration-oriented learning settings (De Jong & Van Joolingen, 1998). Fueled by social cognitive theory, situated cognition, and motivation theories, affective states—conveyed by AI or exhibited in AI-learner interactions—should be tracked and made possible through multimodal computing and machine learning (Guo & Goh, 2015; Lawson et al., 2021; Rickel & Johnson, 1998). To this end, research in AI with affective computing for simulation-based learning can seek progress in two directions. First, pattern recognition and synthesis in machine learning (i.e., affect detection, understanding, and generation). Second, multimodal technologies and cognitive theories of multimedia learning because affect relates to facial expressions, hand and body gestures, voice, and affect classification that should follow evidence-based practice to ensure effective learning.

Multimodal computing and machine learning are also used *for* assessments in simulation-based learning. We identified applicable aspects of AI-powered assessments in simulation-based learning. Sensor technologies were used to collect multimodal data. Benchmark indicators should be developed and used by machine-learning algorithms to establish automatic and meaningful feedback and assessment mechanisms (Castillo-Segura et al., 2021). Given the immersive, action-based, and interactive characteristics of simulation-based learning, we encourage the use of in-simulation AI-driven learning analytics for assessments. We argue that these assessments can be used to provide informative real-time data for teachers to track student learning trajectories; and for students to self-regulate and learn with their best abilities.

4.3. Research and pedagogical implications

AI in simulation-based learning has offered unprecedented opportunities and associated challenges with its interdisciplinary, multidisciplinary, or transdisciplinary characteristics of research (Rosé et al., 2019; Zhang & Aslan, 2021). Our findings shed light on the useful applications of AI in simulation-based learning. We discuss three research and pedagogical implications. First, we encourage the applications of virtual agents in simulation-based learning using AI. AI, adopting NLP, machine learning, or large transformer-based natural language models, can empower virtual agents to provide constructive and personalized learning experience. AI virtual agents can guide and facilitate reflections of learners' decision-making. AIEd researchers can design AI virtual agents to function as peers for collaborative learning or tutees for learners to take an active learning role for teaching and explaining target concepts. Simulation as a learning environment with virtual agents can contextualize and legitimate these learning experiences for knowledge and skills applications. To serve the educational purposes of AI virtual agents, we maintain that AIEd researchers should adopt ensemble and hybrid machine learning techniques with modularized and integrative systems for the representations of domain knowledge and student characteristics as well as the creation of automatic learning assessments. These advanced and integrative approaches are promising in providing naturalistic and authentic discourse as well as facilitating and assessing learners' conceptual understanding and skills development.

In practice, large transformer-based natural language models (Xu et al., 2021) based on deep learning have been developed to be leveraged to solve local educational problems with customization and contextualization. These advanced AI technologies make human-like conversations with virtual agents more authentic and sensible, but they also introduce new challenges. In essence, to appropriately harness the computing power of the AI systems to local educational contexts, we suggest to collect local dataset for the customization and training of the AI systems to adapt to individual contexts for creating a safe and unbiased learning experience for *all* learners (Rosé et al., 2019). Importantly, efforts should be taken to eradicate a lack of participation from minority groups (Cowgill et al., 2020). That said, the local training datasets should be representative as well as inclusive.

Second, we identified multimodality as another unique area of AI in simulation-based learning in which AIEd can observe advancement in the future. Multimodal technologies are distinct in simulation-based learning using AI because of the affordances of multimodal interactions in simulations; we can elaborate from the following two perspectives. From the interface perspective, multimodality in AI for simulation-based learning is related to affective computing. For example, AI-powered virtual agents can demonstrate affective states with verbal, voice, text, or facial expressions. AI can also be used for the detections of learners' affective states with facial, voice, or textual inputs. To design effective simulation-based learning using AI, we recommend that interdisciplinary AIEd research teams to integrate learning principles and theories (e.g., motivation theories, social cognitive theory, and cognitive theories of multimedia learning, see Figure 7) into AI technologies (affective computing with multimodality).

From the embodied perspective, multimodality can afford assessments in simulation-based learning using AI. Embodied assessments have been adopted widely for clinical skills training in medical education. We urge AIEd researchers to design and develop AI assessments with multimodality in simulation-based learning. In-simulation multimodal data logging and mining can be used for creating assessment indicators as the basis of machine-learning-driven mechanisms. The mechanisms can be useful to provide immediate feedback, formative and summative assessments for learners' skills development in simulation-based learning. Teachers should also learn how to use and interpret the results of the assessments for informed decision-making to support effective teaching. We suggest to provide professional development opportunities for teachers to fruitfully integrate AI simulation-based learning environments into their teaching practices.

Finally, we propose that AI in simulation-based learning is interlinked and connected among essential areas such as virtual agents, affective computing, and automatic assessments. AIEd researchers and practitioners should aim to design and develop integrated and coherent systems that promote adaptive and personalized learning with various components. It requires collaborative work among learning designers, educational psychologists, computer programmers, data scientists, and other stakeholders.

Yet, numerous challenges can inhibit the development of holistic simulation-based learning systems for *all* learners. We highlight small datasets in educational research and ethical issues as the two distinct challenges as these were mentioned frequently in our included studies. First, the use of small datasets is a challenge because naturalistic interactions, authentic affective states detections/expressions, and accurate assessments afforded by AI all depend on large input datasets. To contextualize and localize AI technologies in simulation-based learning settings, one needs to use context-specific datasets and train AI with machine learning and deep learning algorithms. Context-specific datasets may not be large enough for accurate machine/deep learning and to build decent digital human learning environments. The innovative development of algorithms and techniques can potentially address this issue. Second, AI ethics is a challenge because, for instance, context-specific data collection with underage students may be protected. AIEd researchers need to abide by ethical practices. This points to the critical needs for the research of AI ethics guidelines creation in educational settings.

5. Limitations

Several limitations should be noted. First, the findings drawn are based on the studies found in the databases and keywords used, thus generalizing the findings for AIEd in simulation-based learning should be cautious. For example, we did not include the Scopus database due to the authors' lack of access. Second, although the coding processes were completed and validated carefully by the authors, they were still human work in which bias may arise (Chen et al., 2020). Finally, we did not exclude repetitive interventions in different studies because each study has their unique contributions and implications, including them serves the purpose of discovering trends in the topic area. However, this practice may cause skewed results. The interpretation of the results should take these limitations into consideration.

6. Conclusion

Following our definitions of AI and simulation-based learning, we bring forward six mapping categories and three interrelated thematic trends for AI in simulation-based learning research. With the promising possibilities given by the growing power of AI computing technologies, we encourage AIEd researchers and practitioners to design with guiding learning principles and theories. We also propose concerns and challenges in the hope to promote learning for *all* learners. Learning should be a balanced triad of teachers/learning designers, learners, and technologies. AI in simulation-based learning aims at pushing the boundary of technologies in service of learner-centered (Graesser & McNamara, 2010; Nussbaumer et al., 2015) and teacher/designer-orchestrated (Holstein et al., 2019) meaningful learning. The benefits of AI in simulation-based learning are mutual between teachers and students (e.g., Hwang et al., 2020): learners can have the luxury to enjoy an equilibrium of challenging tasks and balanced affective states with AI-powered virtual agents, while teachers can take advantages of AI-leveraged assessments to diagnose students' learning trajectories and further improve their teaching.

This review identifies five specific areas in AI-powered simulation-based learning where more research is needed. First, there is a need to continue the line of research in using module-based AI and hybrid AI mechanisms for learning. This line of research can benefit from design-oriented, evaluative, and experimental research designs. Second, researchers should examine the effects of various designs and scaffoldings with the use of AI-powered virtual agents guided by learning principles and theories. Third, researchers should also investigate delayed effects of affective states in AI-powered virtual agents (e.g., in delayed tests). Fourth, AIEd research stakeholders should develop machine/deep learning algorithms or techniques for customization, localization, contextualization with small datasets using unbiased and inclusive approaches to meet the learning needs. Fifth, AIEd research stakeholders should also study ethics-related endeavors in AI in simulation-based learning.

Appendix

Table A1 presented the details of included studies in this systematic review.

Table A1.

Studies and the coding sheet

ID	Included Study	Themes categorized	Methods	AI technologies	Simulation [Learning principles or theories]	Learning outcomes	Research issues and contributions
1	Terzidou et al. (2016)	Virtual agents	Architecture design, experiments, and log files analysis	Natural Language Processing (NLP) (autonomous chats), scripted system	Virtual reality (*OpenSim*) [Motivation, collaborative learning]	The intervention supports collaborative learning, but no effects on learning attitudes.	Improving classroom/community cohesiveness. Time limits virtual agent usefulness.
2	Van der Meij et al. (2015)	Virtual agents	Experimental design	Text to speech (TTS) (*Elckerlyc* and *Loquendo*) and scripted system	Web-based simulation (*SimQuest*) [Motivation, social learning and social constructivism]	There is a significant effect on self-efficacy using virtual agents.	The use of motivational virtual agents is a promising area of research but the design of agents should be carefully considered.
3	Biswas et al. (2005)	Virtual agents	Architecture design and experiments	Module-based modeling architecture: *Betty's Brain*	Web-based simulation [Motivation, social learning and social constructivism, self-regulated learning]	Adding self-regulated support and mentors to better help students to learn new concepts via learning by teaching.	Virtual agents used for learning by teaching. Pursuing naturalistic interactions by using concept representations and reasoning mechanisms.
4	Chin et al. (2010)	Virtual agents	Between subject and crossover experimental design	A qualitative reasoning engine with path traversal algorithms	Web-based simulation [Social learning and social constructivism, self-regulated learning]	Teachable agents' added value (within the intervention and to the curriculum) and transfer effect were confirmed.	Considering added-value and basic-value learning when using educational technology. Preparing students for how to learn or mapping artificial intelligence (AI) technologies with curricular standards.
5	Veletsianos & Russell (2013)	Virtual agents	A basic interpretive study	TTS, the Artificial Intelligence Markup Language (AIML)	Virtual world and simulation games [social learning and social constructivism]	Not applicable	Revealing the nature and diversity of open-ended dialogues and interactions between learners and virtual agents.
6	Kappagantula et al. (2020)	Virtual agents	User study and expert evaluation	Module-based automatic system for script generation	Web-based simulation [Motivation, situated cognition]	Not applicable	The design of a system that automatically generates deictic gestures for virtual agents.
7	Nurshatayeva et al. (2021)	Virtual agents	Randomized Controlled Trial	Machine learning chatbot: *PeeDee*	Campus-specific web-based chatbot [Personalized/self-regulated learning]	AI chatbot is effective for college freshmen to navigate pre-enrollment requirements. Improving college access.	An example of effective low-cost chatbot intervention for college matriculation.
8	Ieronutti & Chittaro (2007)	Virtual agents	Conceptual (architecture design) and case studies	Module-based automatic system	Web-based simulation game [Social learning and social constructivism]	Not applicable	Presenting architecture for compatible interactive virtual agents integration in web-based simulations.
9	Nye et al. (2021)	Virtual agents	Architecture design and a usability study	Ensemble methods in machine learning and NLP	Web-based simulation [Social learning and social constructivism, motivation]	Not applicable	Multi-agent pedagogy, cold-start mitigation, generalist vs. specialist agent role, and mentor career coverage.
10	Johnson (2003a)	Virtual agents	Conceptual paper (architecture design)	NLP	Web based simulation training system [Social cognitive learning, situated cognition]	Not applicable	Providing interactive tutorial dialog and interaction tactics design examples.
11	Matsuda et al. (2020)	Virtual agents	Architecture design and experiments	Module-based modeling architecture: *Betty's Brain*	Intelligent Tutoring Systems (ITSs) [Motivation (limited), self-regulated learning]	Learning-by-teaching agents increase equation solving proficiency; metacognitive scaffolding is equally effective regardless of students' prior competency.	Suggesting naturalistic interactions and constructive and reflective questions prompted by the agents.
12	Rickel & Johnson, (1998)	Virtual agents	Conceptual (architecture design)	Module-based modeling architecture: cognitive components, sensorimotor processing, and speech recognition	ITSs (*Steve*) [Motivation (attention), social learning and social constructivism]	Not applicable	Steve, a human-like agent is the result of interdisciplinary research to interact with and help students to learn.
13	Kumar & Rosé (2011)	Virtual agents	Evaluation with Experiments	Module-based modeling architecture called *Basilica*	Second life virtual world [Collaborative learning]	The most recent agents designed with social interaction strategies were effective (0.71σ) compared to earlier versions.	Multi-agents using *Basilica* provided a rich representational capability. An example of systematic, distributable, and efficient design.
14	Schroeder et al. (2013)	Virtual agents	Meta-analysis	AI without technical details	ITSs [Social learning, cognitive theory of multimedia learning]	Agents are more beneficial in K-12 than post-secondary contexts; on-screen text is better for learning than narration.	Suggesting detailed reporting and designs for AI agent studies; and more experimental studies on K-12 education.
15	Kim & Baylor (2016)	Virtual agents/Affects	Conceptual paper	Rule-based system/real-time adaptive feedback	Web-based simulation [Social learning and social constructivism, motivation]	Not applicable	Highlighting the importance of emotions in multi-agents and instructional roles in agent studies.

16	Johnson & Lester (2016)	Virtual agents/Affects	Conceptual paper	NLP, affective computing	Interactive simulation microworld [Social learning and social constructivism, motivation]	Not applicable	Various uses of virtual agents were outlined with preliminary favorable evidence, indicating future research is needed.
17	D'Mello & Graesser (2012)	Virtual agents/Affects	Architecture design and evaluation with experiments	Module-based automatic system: Latent Semantic Analysis, affective computing: Affective *AutoTutor*	ITSs [Social cognitive learning, motivation]	*Affective AutoTutor* demonstrated remarkable learning improvements compared to the original *AutoTutor*.	Various *AutorTutor*-derived systems were introduced. AIEd scholars can make reference to.
18	Lawson et al. (2021)	Virtual Agents/Affects	2 by 2 between subject design	Rule-based AI system	Web-based simulation [Positivity principle of the cognitive theory of multimedia learning]	Learners perceived virtual agents with positive emotions to be better for learning and engagement; but no effects for learning on a delayed test.	The design of online learning experiences with virtual agents showing emotions. Revealing the gap for more research on the positivity principle for improved learning.
19	Guo & Goh (2016)	Virtual agents/Affects	Between subject experimental design	Scripted (minimal) AI	Web-based simulation game [Social learning and social constructivism, motivation]	The authors did not find a significant difference in learning outcomes between affective and nonaffective agent conditions; the difference on *motivation* was significant.	Promoting the use of AI agents to enhance motivation. Suggesting iterative design studies to investigate design tradeoffs.
20	Gulz & Haake (2006)	Virtual agents/Affects	Conceptual (review)	Module-based modelling architecture: domain knowledge, student model and teaching knowledge	Simulation games [Social learning and social constructivism, motivation, self-regulated learning]	Not applicable	A comprehensive review on virtual agents' look that influences affective states and learning.
21	Lippert et al. (2020)	Virtual agents/Affects	Conceptual paper (architecture design)	NLP	ITSs [Social cognitive theory, motivation]	Not applicable	Suggesting that research in conversational ITSs with multi-agents should specify how learner emotional states are influenced by designs.
22	Guo & Goh (2015)	Virtual agents/Affects	Meta-analysis	AI without technical details	Computer-based simulated learning environment [Social cognitive learning, motivation]	Infusing emotions in virtual agents has significant effects on learners' motivation to learn, knowledge retention, and transfer.	Conceptual studies are needed to dissect the elements useful for impacting learners' affective states with affective computing using AI agents.
23	Craig & Schroeder (2017)	Virtual agents/Affects	A randomized alternative treatments design with a pretest	TTS	ITSs-paced video [Cognitive theory of multimedia learning]	The voice effect was tested favoring modern TTS engine voice (i.e., significant learning transfer and training efficiency).	Voice effects of AI agents on learning and learners' perceptions. (Dis) matching modern-machine voice or human voice with agent appearances.
24	Guran et al. (2020)	Affects/Assessme nts	Architecture design and machine learning validation	Convolution Neural Networks	Smart edutainment [Motivation, social learning and social constructivism]	Not applicable	Presenting challenges in the design steps, in the preschooler context.
25	Clancey & Hoffman (2021)	Affects/Assessme nts	Conceptual paper	Module-based system with machine learning (deep neural networks)	ITSs [Situated cognition, social cognitive learning]	Not applicable	Mutually informing both Explainable Artificial Intelligence (XAI) and ITSs to help human problem-solving.
26	Troussas et al. (2020)	Affects/Assessme nts	Evaluation with Experiments	Module-based automatic system/fuzzy-modeled personalization	Mobile simulation game-based learning [Personalized/self-regulated learning, collaborative learning]	Cognitive advice and recommendation for peer collaboration were effective for knowledge and skills development.	The results can be used as guidelines for pedagogically sound and pleasant learning.
27	Hooshyar et al. (2020)	Affects/Assessme nts	A systematic literature review	Module-based modeling architecture: Open Learner Modelling (OLM)	Included articles using virtual simulations [Motivation, self-regulated learning]	Not applicable	A comprehensive review on open learner models including strategies, and how such models support cognition, metacognition, and motivation. Emotional supports were rarely used in open learner models.
28	Moon et al. (2020)	Affects/Assessme nts	Machine learning validation	NLP and machine learning (support vector machine (SVM), Naïve Bayes, Decision trees, logistic regression)	Virtual reality (*OpenSim*) [Social cognitive theory]	Not applicable	The feasibility of automatic assessments of cognitive and emotional states with machine learning for children with autism to provide adaptive support.
29	Jaques & Vicari (2007)	Affects/Assessme nts	Conceptual paper (architecture design)	Module-based automatic system/affective computing: Belief-Desire-Intention (BDI) model	Web-based simulation [Situated cognition, motivation]	Not applicable	The BDI model is dynamic for modeling, capturing, and deducing students' affective states.
30	Sarrafzadeh et al. (2008)	Affects/Assessme nts	Conceptual paper (architecture design)	Module-based automatic system, affective computing/fuzzy approach	ITSs [Motivation, social cognitive theory]	Not applicable	The AI system was able to capture students' cognitive and affective states and provide adaptive learning experience.
31	Aleven et al. (2017)	Affects/Assessme nts	Conceptual (review)	Adaptive technologies and algorithms (e.g., Bayesian Networks and reinforcement learning)	ITSs [Self-regulated learning]	Not applicable	Proposed three forms of adaptivity (p.6), and five broad psychological realms (pp. 49-50). Hybrid forms of adaptivity are promising.

32	Rhienmora et al. (2011)	Assessments	Experiment/machine learning validation	Automatic feedback and assessment (Hidden Markov Models, HMMs), haptic (multimodal) technologies	Medical (dental) training simulations [Social learning and social constructivism (feedback)]	Experiments suggested that the dental skills assessment using HMMs is accurate.	Providing practical examples of surgical (dental) technical skills assessments.
33	Peters et al. (2021)	Assessments	Survey design with tests and log data for analysis	Supervised machine learning (random forest regression models and a nested cross validation scheme)	Minecraft simulation game [Motivation]	Not applicable (non-experimental design, tests were used for structural equation modelling and machine learning)	Log data with machine learning are predictive of external tests of performance. The authors suggested that game-based intelligence assessment is appropriate using Minecraft.
34	Amershi & Conati (2009)	Assessments	Architecture design and evaluation with experiments	Module-based system with unsupervised and supervised machine learning clustering	Exploratory Learning Environments (interactive simulations) [Social cognitive theory, self-regulated learning]	Not applicable (experiments were for modeling evaluations)	A data-based framework for user modeling that uses both unsupervised and supervised classification. May be used for adaptive learning support development.
35	Nussbaumer et al. (2015)	Assessments	Architecture design and descriptive evaluation study design	Module-based modeling architecture: OLM	Web-based virtual environment [Motivation, self-regulated learning]	Not applicable	Promising results evaluating module-based modelling architecture, promoting the use of OLM.
36	Gandedkar et al. (2021)	Assessments	Literature review	Module-based modeling architecture: OLM	ITSs, virtual reality [Situated cognition, self-regulated learning]	Not applicable	Suggesting dynamic engagement of students/educators in a “pedagogical, ethically-appropriate, community-centered, culturally-sensitive and economically-feasible milieu” (p. 75).
37	Gardner et al. (2021)	Assessments	Conceptual paper (review)	Machine learning for automated essay scoring systems and computerized adaptive tests	Computer adaptive testing in online simulation [Self-regulated learning]	Not applicable	Technological advancement is in the analysis of large-scale assessment process data; but the percepts and functions of educational assessment have not changed too much.
38	Castillo-Segura et al. (2021)	Assessments	Systematic literature review	Internet of Things (for multimodal sensors) and machine learning	Medical training simulations [Social learning (feedback), motivation, self-regulated learning]	Not applicable	Providing comprehensive review of sensors, indicators for assessments, and algorithms for surgical technical skills assessments.
39	Katz et al. (2021)	Assessments	Architecture design and a pilot experiment	Module-based modeling architecture: machine learning clustering, adaptive algorithms, natural language recognition	ITSs (*Rimac*) [Social learning (ZPD)]	Student modeling-driven tutorial dialogue systems improved learning, but not better than the control group (non-dynamic version).	Suggesting the adoption of OLM, student modelling for effective and dynamic ITSs.
40	Yang et al. (2019)	Assessments	Machine learning validation	Artificial neural networks for assessments	Web-based simulation [Cognitive learning]	Not applicable	The results of machine learning assessments can inform teachers and the design of interventions on L2 acquisition.
41	Westera et al. (2018)	Assessments	Machine learning validation	NLP with *ReaderBench* framework	Web-based simulation game [Social learning, self-regulated learning]	Not applicable	On essay scoring: applications for learners' self-assessment/evaluation and reductions of teacher load.
42	Lin et al. (2013)	Assessments	Experimental design	Module-based system with machine learning (decision trees) for personalization; research: rule-based agent system	Web-based simulation game [Personalized/self-regulated learning, motivation]	Learners have a 90% probability to achieve above-average creativity score if an AI suggested optimal learning path is used.	Decision trees provide a personalized creativity learning system. The system should consider the learning behavior, the physical environment, the personal traits, and the teacher behavior.
43	Legaspi et al. (2008)	Assessments	Evaluation with Experiments	Module-based predictive system, unsupervised cluster modeling	ITSs [Situated cognition]	Not applicable (analyses of the experiments were correlational)	Constructed predictive models of high-level student information that can support personalized adaptation.
44	Winkler-Schwartz et al. (2019)	Assessments	Systematic literature review	Machine learning	Virtual reality simulation [Situated cognition, social learning and social constructivism]	Not applicable	On the applications of machine learning to assess psychomotor performance in medicine using virtual reality simulators. Provided a checklist for this interdisciplinary area.
45	Towers et al. (2019)	Assessments	Scoping review	Automatic feedback and scoring, haptic (multimodal) technologies	Virtual reality simulation [Social learning and social constructivism (feedback)]	Not applicable	Suggesting a lack of educational standards and associated exercises for dental simulators. The scoring mechanisms have not been validated with actual clinical performance.
46	Molenaar et al. (2021)	Assessments	A pretest/posttest design	Intervention: Adaptive algorithm; research: Bayesian Knowledge Tracing (BKT)	ITSs [Self-regulated learning]	A significant improvement on problem-solving learning was found with a medium effect size.	Moment-by-moment learning curves indicate accuracy regulation in adaptive learning. May be used to support learning with personalized visualizations

47	Forbus et al. (1999)	Assessments	Conceptual (architecture design)	Module-based system with a combination of teleological and analogical/case-based reasoning	Virtual laboratory (*CyclePad*) [Situated cognition (coaching/articulation)]	Not applicable	Simulated lab with AI is promising to provide learner support. The authors concluded with several technical innovations and surprises.
48	Johnson (2003b)	Assessments/Virtual agents	Conceptual paper (architecture design)	Module-based modeling architecture, hypothesis-based reasoning	Web based simulation training system [Social cognitive learning]	Not applicable	Detailed architecture of virtual agents design; suggesting future research on promoting social interactions between humans and agents and rapid design approach.
49	Graesser et al. (2008)	Assessments/Virtual agents	Conceptual paper (architecture design)	NLP	Interactive simulation microworld [Social cognitive learning, self-regulated learning]	Not applicable	Introducing various virtual systems used for effective knowledge construction, highlighting the role of prior knowledge
50	Song et al. (2004)	Assessments/Virtual agents	Conceptual paper (architecture design)	Module-based automatic system (latent semantic analysis): *AutorTutor*	Web-based simulation [Social learning and social constructivism]	Not applicable	Rapid and cost-effective development of AI-powered conversational agents is possible with *AutorTutor* system.
51	Graesser et al. (2017)	Assessments/Virtual agents	Conceptual paper (architecture design)	Module-based automatic system (latent Semantic Analysis): *AutorTutor*	ITSs [Social learning, collaborative learning, motivation]	Not applicable	Conversation-based assessments (CBA) with multi-agents (playing roles of tutor and peer). Advocating that theory is needed for CBA.
52	Graesser & McNamara (2010)	Assessments/Virtual agents	Conceptual (ITSs review)	AI without technical details	ITSs [Self-regulated learning]	Not applicable	Suggesting discourse move and fine-grained adaptivity with intelligent agents to promote self-regulated learning.
53	de Freitas & Neumann (2009)	Assessments/Virtual agents	Conceptual paper with case studies	Automated feedback: advanced AI (no technical details)	Immersive game-based simulation [Situated cognition, self-regulated learning]	Case studies suggested that immersive learning can support "learning changes," but the results needed to be confirmed with experimental design.	Proposing exploratory learning model for simulation game-based learning using advanced AI.
54	Nye et al. (2014)	Assessments/Virtual agents	Conceptual paper (architecture design)	Natural language tutoring, latent semantic analysis (semantic matching algorithm)	Interactive simulation microworld [Social cognitive learning, self-regulated learning, motivation]	Not applicable	Expectation- and misconception-tailored dialogue. Suggesting to empirically test the combinations of virtual agents' tutoring features (see Table 4 in Nye et al. (2014), p. 456 for more information).
55	Dalinger et al. (2020)	Assessments/Virtual agents	An instrumental case study	A blend of AI and live actors	Mixed reality simulation [Situated cognition, self-regulated learning, motivation (confidence)]	Four qualitative themes: opportunity for authentic practice, perceived transfer of learning, perceived confidence, and challenges of using the mixed reality simulation.	Implications for teacher preparation using mixed reality simulations with artificial agents.
56	Ke et al. (2016)	Virtual agents	A mixed methods case study	A mix of virtual agents controlled by the AI script and student avatars played by peer trainees. Multimodal inputs	A mixed-reality integrated learning environment [Social learning and social constructivism, situated cognition, motivation]	Mixed reality supports the sense of presence and the teaching performance of college teaching assistants with various virtual tasks/actions, avatar-embodied live gesturing, environmental fidelity.	Using emerging sensory technologies for AI in simulation-based learning and exploring how design features afford situated learning and transfer.
57	Lorenzo et al. (2013)	Assessments/Virtual agents	A mixed methods study	AI: adaptive agents for student inputs detection, multimodal assessments	Immersive virtual learning environments [Motivation, social learning and social constructivism]	Increased understanding and better task performance were observed for students with Asperger syndrome.	An integrative approach/protocol helped students with Asperger syndrome in simulation-based learning.
58	Latham et al. (2014)	Assessments/Virtual agents	Experimental design	Adaptive algorithms and the Felder–Silverman engineering learning styles model	ITSs [Social cognitive theory, motivation]	Students used adaptive tutorials and matched to their learning profiles outperformed those who used unmatched one.	The adaptive algorithms driving personalized natural language tutorial are generalizable. Adapting to students' engineering learning states can improve learning.
59	Flores et al. (2013)	Assessments/Virtual agents	Architecture design and an evaluation	Module-based system with Bayesian Networks	Multiagent virtual simulation (*The SimDeCS)* [Situated cognition, motivation]	Not applicable (participants' self-report evaluations)	Bayesian networks are helpful for medical learning (knowledge and reasoning).

Reference

***denotes included studies in the systematic review**

*Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2017). Instruction based on adaptive learning technologies. In R. E. Mayer, & P. Alexander (Eds.), *Handbook of research on learning and instruction* (2nd ed., pp. 522–560). New York: Routledge.

Alinier, G. (2007). A typology of educationally focused medical simulation tools. *Medical teacher, 29*(8), e243-e250. https://doi.org/10.1080/01421590701551185

*Amershi, S., & Conati, C. (2009). Combining unsupervised and supervised classification to build user models for exploratory learning environments. *Journal of educational data mining, 1*(1), 18-71. https://doi.org/10.5281/zenodo.3554659

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice Hall.

*Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & The Teachable Agents Group at Vanderbilt. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence, 19*(3-4), 363-392. https://doi.org/10.1080/08839510590910200

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint.* https://arxiv.org/abs/2005.14165

*Castillo-Segura, P., Fernández-Panadero, C., Alario-Hoyos, C., Muñoz-Merino, P. J., & Kloos, C. D. (2021). Objective and automated assessment of surgical technical skills with IoT systems: A systematic literature review. *Artificial Intelligence in Medicine, 112*, 102007, 1-17. https://doi.org/10.1016/j.artmed.2020.102007

Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory gaps during the rise of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence, 1,* 100002, 1-20. https://doi.org/10.1016/j.caeai.2020.100002

Chen, X., Zou, D., Cheng, G., & Xie, H. (2020). Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of Computers & Education. *Computers & Education, 151*, 103855, 1-21. https://doi.org/10.1016/j.compedu.2020.103855

*Chin, D. B., Dohmen, I. M., Cheng, B. H., Oppezzo, M. A., Chase, C. C., & Schwartz, D. L. (2010). Preparing students for future learning with teachable agents. *Educational Technology Research and Development, 58*(6), 649-669. https://doi.org/10.1007/s11423-010-9154-5

Choi, J. I., & Hannafin, M. (1995). Situated cognition and learning environments: Roles, structures, and implications for design. *Educational technology research and development, 43(*2), 53-69. https://doi.org/10.1007/BF02300472

*Clancey, W. J., & Hoffman, R. R. (2021). Methods and standards for research on explainable artificial intelligence: Lessons from intelligent tutoring systems. *Applied AI Letters*, *2*(4). 1-8. https://doi.org/10.1002/ail2.53

Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N., & Chaintreau, A. (2020). Biased programmers? or biased data? a field experiment in operationalizing AI ethics. In *Proceedings of the 21st ACM Conference on Economics and Computation* (pp. 679-681).

*Craig, S. D., & Schroeder, N. L. (2017). Reconsidering the voice effect when learning from a virtual human. *Computers & Education, 114*, 193-205. https://doi.org/10.1016/j.compedu.2017.07.003

Dai, C-P., Ke, F., Dai, Z., West, L., Bhowmik, S., & Yuan, X. (2021). Designing Artificial Intelligence (AI) in Virtual Humans for Simulation-Based Training with Graduate Teaching Assistants. In de Vries, E., Hod, Y., & Ahn J. (Eds.). In *Proceedings of the 15th International Conference of the Learning Sciences - ICLS 2021* (pp. 1101-1102). Bochum, Germany: International Society of the Learning Sciences. https://repository.isls.org/bitstream/1/7418/1/1101-1102.pdf

*Dalinger, T., Thomas, K. B., Stansberry, S., & Xiu, Y. (2020). A mixed reality simulation offers strategic practice for pre-service teachers. *Computers & Education, 144*, 103696, 1-15. https://doi.org/10.1016/j.compedu.2019.103696

*de Freitas, S., & Neumann, T. (2009). The use of 'exploratory learning' for supporting immersive learning in virtual environments. *Computers & Education, 52*(2), 343-352. https://doi.org/10.1016/j.compedu.2008.09.010

De Jong, T., & Van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. Review of educational research, 68(2), 179-201. https://doi.org/10.3102/00346543068002179

*D'Mello, S., & Graesser, A. (2012). AutoTutor and affective autotutor. *ACM Transactions on Interactive Intelligent Systems, 2*(4), 1-39. http://dx.doi.org/10.1145/2395123.2395128.

*Flores, C. D., Barros, P., Cazella, S., & Bez, M. R. (2013). Leveraging the learning process in health through clinical cases simulator. In *2013 IEEE 2nd International Conference on Serious Games and Applications for Health (SeGAH)* (pp. 1-6). IEEE. https://doi.org/10.1109/SeGAH.2013.6665303

*Forbus, K. D., Kuehne, S. E., Whalley, P. B., Everett, J. O., Ureel, L., Brokowski, M., et al. (1999). Cyclepad: an articulate virtual laboratory for engineering thermos dynamics. *Artificial Intelligence, 114*(1-2), 297-347. https://doi.org/10.1016/S0004-3702(99)00080-6

*Gandedkar, N. H., Wong, M. T., & Darendeliler, M. A. (2021). Role of Virtual Reality (VR), Augmented Reality (AR) and Artificial Intelligence (AI) in tertiary education and research of orthodontics: An insight. *Seminars in Orthodontics, 27*(2), 69-77. https://doi.org/10.1053/j.sodo.2021.05.003

*Gardner, J., O'Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment:'Breakthrough? Or buncombe and ballyhoo?'. *Journal of Computer Assisted Learning, 37*(5), 1207-1216. https://doi.org/10.1111/jcal.12577

*Graesser, A. C., Cai, Z., Morgan, B., & Wang, L. (2017). Assessment with computer agents that engage in conversational dialogues and trialogues with learners. *Computers in Human Behavior, 76*, 607-616. https://doi.org/10.1016/j.chb.2017.03.041

Graesser, A. C., Conley, M. W., & Olney, A. (2012). Intelligent tutoring systems. In K. R. Harris, S. Graham, T. Urdan, A. G. Bus, S. Major, & H. L. Swanson (Eds.), APA educational psychology handbook, Vol. 3. Application to learning and teaching (pp. 451–473). American Psychological Association. https://doi.org/10.1037/13275-018

*Graesser, A. C., Jeon, M., & Dufty, D. (2008). Agent technologies designed to facilitate interactive knowledge construction. *Discourse processes, 45*(4-5), 298-322. https://doi.org/10.1080/01638530802145395

*Graesser, A., & McNamara, D. (2010). Self-regulated learning in learning environments with pedagogical agents that interact in natural language. *Educational Psychologist, 45*(4), 234-244. https://doi.org/10.1080/00461520.2010.515933

*Gulz, A., & Haake, M. (2006). Design of animated pedagogical agents—A look at their look. *International Journal of Human-Computer Studies, 64*(4), 322-339. https://doi.org/10.1016/j.ijhcs.2005.08.006

*Guo, Y. R., & Goh, D. H. L. (2015). Affect in embodied pedagogical agents: Meta-analytic review. *Journal of Educational Computing Research, 53*(1), 124-149. https://doi.org/10.1177/0735633115588774

*Guo, Y. R., & Goh, D. H. L. (2016). Evaluation of affective embodied agents in an information literacy game. *Computers & Education, 103*, 59-75. https://doi.org/10.1016/j.compedu.2016.09.013

*Guran, A. M., Cojocar, G. S., & Dioşan, L. S. (2020). Developing smart edutainment for preschoolers: A multidisciplinary approach. In *Proceedings of the 2nd ACM SIGSOFT International Workshop on Education through Advanced Software Engineering and Artificial Intelligence* (pp. 20-26). https://doi.org/10.1145/3412453.3423197

Heradio, R., De La Torre, L., Galan, D., Cabrerizo, F. J., Herrera-Viedma, E., & Dormido, S. (2016). Virtual and remote labs in education: A bibliometric analysis. *Computers & Education, 98*, 14-38. https://doi.org/10.1016/j.compedu.2016.03.010

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Designing for complementarity: Teacher and student needs for orchestration support in AI-enhanced classrooms. In *International Conference on Artificial Intelligence in Education* (pp. 157-171). Springer, Cham. https://doi.org/10.1007/978-3-030-23204-7_14

*Hooshyar, D., Pedaste, M., Saks, K., Leijen, Ä., Bardone, E., & Wang, M. (2020). Open learner models in supporting self-regulated learning in higher education: A systematic literature review. *Computers & Education, 154*, 103878, 1-19. https://doi.org/10.1016/j.compedu.2020.103878

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence, 1*, 100001, 1-5. https://doi.org/10.1016/j.caeai.2020.100001

*Ieronutti, L., & Chittaro, L. (2007). Employing virtual humans for education and training in X3D/VRML worlds. *Computers & Education, 49*(1), 93-109. https://doi.org/10.1016/j.compedu.2005.06.007

*Jaques, P. A., & Vicari, R. M. (2007). A BDI approach to infer student's emotions in an intelligent learning environment. *Computers & Education, 49*(2), 360-384. https://doi.org/10.1016/j.compedu.2005.09.002

*Johnson, W. L. (2003a). Interaction tactics for socially intelligent pedagogical agents. In *Proceedings of the 8th international conference on Intelligent user interfaces* (pp. 251-253). https://doi.org/10.1145/604045.604090

*Johnson, W. L. (2003b). Using agent technology to improve the quality of web-based education. In *Web Intelligence* (pp. 77-101). Springer.

*Johnson, W. L., & Lester, J. C. (2016). Face-to-face interaction with pedagogical agents, twenty years later. *International Journal of Artificial intelligence in education, 26*(1), 25-36. https://doi.org/10.1007/s40593-015-0065-9

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

*Kappagantula, S. R. K., Adamo-Villani, N., Wu, M. L., & Popescu, V. (2020). Automatic deictic gestures for animated pedagogical agents. *IEEE Transactions on Learning Technologies, 13*(1), 1-13. https://doi.org/10.1109/TLT.2019.2922134

*Katz, S., Albacete, P., Chounta, I. A., Jordan, P., McLaren, B. M., & Zapata-Rivera, D. (2021). Linking dialogue with student modelling to create an adaptive tutoring system for conceptual physics. *International Journal of Artificial Intelligence in Education, 31*, 397-445. https://doi.org/10.1007/s40593-020-00226-y

*Ke, F., Lee, S., & Xu, X. (2016). Teaching training in a mixed-reality integrated learning environment. *Computers in Human Behavior, 62*, 212-220. https://doi.org/10.1016/j.chb.2016.03.094

Keller, J. M. (2010). *Motivational design for learning and performance. The ARCS model approach*. Springer Verlag.

*Kim, Y., & Baylor, A. L. (2016). Research-based design of pedagogical agent roles: A review, progress, and recommendations. *International Journal of Artificial Intelligence in Education, 26*(1), 160-169. https://doi.org/10.1007/s40593-015-0055-y

*Kumar, R., & Rosé, C. P. (2011). Architecture for building conversational agents that support collaborative learning. *IEEE Transactions on Learning Technologies, 4*(1), 21-34. https://doi.org/10.1109/TLT.2010.41

Kunkler, K. (2006). The role of medical simulation: an overview. *The International Journal of Medical Robotics and Computer Assisted Surgery, 2*(3), 203-210. https://doi.org/10.1002/rcs.101

*Latham, A., Crockett, K., & McLean, D. (2014). An adaptation algorithm for an intelligent natural language tutoring system. *Computers & Education, 71*, 97-110. https://doi.org/10.1016/j.compedu.2013.09.014

*Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021). Do Learners Recognize and Relate to the Emotions Displayed By Virtual Instructors?. *International Journal of Artificial Intelligence in Education, 31*(1), 134-153. https://doi.org/10.1007/s40593-021-00238-2

Leahy, S. M., Holland, C., & Ward, F. (2019). The digital frontier: Envisioning future technologies impact on the classroom. *Futures, 113*, 102422, 1-10. https://doi.org/10.1016/j.futures.2019.04.009

*Legaspi, R., Sison, R., Fukui, K. I., & Numao, M. (2008). Cluster-based predictive modeling to improve pedagogic reasoning. *Computers in Human Behavior, 24*(2), 153-172. https://doi.org/10.1016/j.chb.2007.01.007

*Lin, C. F., Yeh, Y. C., Hung, Y. H., & Chang, R. I. (2013). Data mining for providing a personalized learning path in creativity: An application of decision trees. *Computers & Education, 68*, 199-210. https://doi.org/10.1016/j.compedu.2013.05.009

*Lippert, A., Shubeck, K., Morgan, B., Hampton, A., & Graesser, A. (2020). Multiple agent designs in conversational intelligent tutoring systems. *Technology, Knowledge and Learning, 25*, 443-463. https://doi.org/10.1007/s10758-019-09431-8

Loder, J., Nicholas, L. (2018). *Confronting Dr Robot: Creating a people-powered future, for AI in health*. Retrieved from: https://media.nesta.org.uk/documents/confronting_dr_robot.pdf

*Lorenzo, G., Pomares, J., & Lledó, A. (2013). Inclusion of immersive virtual learning environments and visual control systems to support the learning of students with Asperger syndrome. *Computers & Education, 62*, 88-101. https://doi.org/10.1016/j.compedu.2012.10.028

Luckin, R., & Cukurova, M. (2019). Designing educational technologies in the age of AI: A learning sciences-driven approach. *British Journal of Educational Technology, 50*(6), 2824-2838. https://doi.org/10.1111/bjet.12861

Maroengsit, W., Piyakulpinyo, T., Phonyiam, K., Pongnumkul, S., Chaovalit, P., & Theeramunkong, T. (2019). A survey on evaluation methods for chatbots. In *Proceedings of the 2019 7th International Conference on Information and Education Technology* (pp. 111-119). https://doi.org/10.1145/3323771.3323824

*Matsuda, N., Weng, W., & Wall, N. (2020). The effect of metacognitive scaffolding for learning by teaching a teachable agent. *International Journal of Artificial Intelligence in Education, 30*, 1-37. https://doi.org/10.1007/s40593-019-00190-2

Mayer, R. E. (2020). *Multimedia learning* (3rd ed.). Cambridge University Press.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI magazine, 27*(4), 12-14. https://doi.org/10.1609/aimag.v27i4.1904

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education, 70*, 29-40. https://doi.org/10.1016/j.compedu.2013.07.033

Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). *Qualitative data analysis: A methods sourcebook* (4th ed.). Sage Publications.

Miller, M. D. (1984). The use of simulations in training programs: A review. *Educational technology, 24(*11), 39-42. https://www.jstor.org/stable/44424271

*Molenaar, I., Horvers, A., & Baker, R. S. (2021). What can moment-by-moment learning curves tell about students' self-regulated learning?. *Learning and Instruction, 72*, 101206, 1-14. https://doi.org/10.1016/j.learninstruc.2019.05.003

*Moon, J., Ke, F., & Sokolikj, Z. (2020). Automatic assessment of cognitive and emotional states in virtual reality-based flexibility training for four adolescents with autism. *British Journal of Educational Technology, 51*(5), 1766-1784. https://doi.org/10.1111/bjet.13005

Myllyaho, L., Raatikainen, M., Männistö, T., Mikkonen, T., & Nurminen, J. K. (2021). Systematic literature review of validation methods for AI systems. *Journal of Systems and Software, 181*, 111050, 1-22. https://doi.org/10.1016/j.jss.2021.111050

*Nurshatayeva, A., Page, L.C., White, C.C. et al. (2021). Are Artificially Intelligent Conversational Chatbots Uniformly Effective in Reducing Summer Melt? Evidence from a Randomized Controlled Trial. *Research in Higher Education, 62*, 392-402. https://doi.org/10.1007/s11162-021-09633-z

*Nussbaumer, A., Hillemann, E.-C., Gütl, C., & Albert, D. (2015). A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments. *Journal of Learning Analytics, 2*(1), 101–133. https://doi.org/10.18608/jla.2015.21.6

*Nye, B. D., Davis, D. M., Rizvi, S. Z., Carr, K., Swartout, W., Thacker, R., & Shaw, K. (2021). Feasibility and usability of MentorPal, a framework for rapid development of virtual mentors. *Journal of Research on Technology in Education, 53*(1), 21-43. https://doi.org/10.1080/15391523.2020.1771640

*Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education, 24*(4), 427-469. https://doi.org/10.1007/s40593-014-0029-5

Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of artificial intelligence research, 11*, 169-198. https://doi.org/10.1613/jair.614

Pan, Z., Cheok, A. D., Yang, H., Zhu, J., & Shi, J. (2006). Virtual reality and mixed reality for virtual learning environments. *Computers & graphics, 30*(1), 20-28. https://doi.org/10.1016/j.cag.2005.10.004

Perez, S., Massey-Allard, J., Butler, D., Ives, J., Bonn, D., Yee, N., & Roll, I. (2017). Identifying productive inquiry in virtual labs using sequence mining. In E. André, R. Baker, X. Hu, M. M. T. Rodrigo, & B. du Boulay (Eds.), *Artificial intelligence in education*, (vol. 10, 331, pp. 287–298). https://doi.org/10.1007/978-3-319-61425-0_24.

*Peters, H., Kyngdon, A., & Stillwell, D. (2021). Construction and validation of a game-based intelligence assessment in minecraft. *Computers in Human Behavior, 119*, 106701, 1-13. https://doi.org/10.1016/j.chb.2021.106701

Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and software technology, 64*, 1-18. https://doi.org/10.1016/j.infsof.2015.03.007

Picard, R. W. (1997). *Affective computing*. MIT Press.

*Rhienmora, P., Haddawy, P., Suebnukarn, S., & Dailey, M. N. (2011). Intelligent dental training simulator with objective skill assessment and feedback. *Artificial intelligence in medicine, 52*(2), 115-121. https://doi.org/10.1016/j.artmed.2011.04.003

*Rickel, J., & Johnson, W. L. (1998). STEVE: A pedagogical agent for virtual reality. In *Proceedings of the second international conference on Autonomous agents* (pp. 332-333). https://doi.org/10.1145/280765.280851

Rosé, C. P., McLaughlin, E. A., Liu, R., & Koedinger, K. R. (2019). Explanatory learner models: Why machine learning (alone) is not the answer. *British Journal of Educational Technology, 50*(6), 2943-2958. https://doi.org/10.1111/bjet.12858

*Sarrafzadeh, A., Alexander, S., Dadgostar, F., Fan, C., & Bigdeli, A. (2008). "How do you know that I don't understand?" A look at the future of intelligent tutoring systems. *Computers in Human Behavior, 24*(4), 1342-1363. https://doi.org/10.1016/j.chb.2007.07.008

*Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). How effective are pedagogical agents for learning? A meta-analytic review. *Journal of Educational Computing Research, 49*(1), 1-39. https://doi.org/10.2190/EC.49.1.a

Schunk, D. H., & DiBenedetto, M. K. (2020). Motivation and social cognitive theory. *Contemporary Educational Psychology, 60*, 101832, 1-10. https://doi.org/10.1016/j.cedpsych.2019.101832

Shum, S. J. B., & Luckin, R. (2019). Learning analytics and AI: Politics, pedagogy and practices. *British journal of educational technology, 50*(6), 2785-2793. https://doi.org/10.1111/bjet.12880

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel psychology, 64*(2), 489-528. https://doi.org/10.1111/j.1744-6570.2011.01190.x

*Song, K. S., Hu, X., Olney, A., Graesser, A. C., & Tutoring Research Group. (2004). A framework of synthesizing tutoring conversation capability with web-based distance education courseware. *Computers & Education, 42*(4), 375-388. https://doi.org/10.1016/j.compedu.2003.09.002

Spronck, P., Ponsen, M., Sprinkhuizen-Kuyper, I., & Postma, E. (2006). Adaptive game AI with dynamic scripting. *Machine Learning, 63*(3), 217-248. https://doi.org/10.1007/s10994-006-6205-6

Stiefelhagen, R., Bernardin, K., Bowers, R., Garofolo, J., Mostefa, D., Soundararajan, P. (2007). The CLEAR 2006 Evaluation. In: Stiefelhagen, R., Garofolo, J. (eds). *Multimodal Technologies for Perception of Humans. CLEAR 2006. Lecture Notes in Computer Science*, vol 4122. (pp. 1-44) Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-69568-4_1

*Terzidou, T., Tsiatsos, T., Miliou, C., & Sourvinou, A. (2016). Agent supported serious game environment. *IEEE Transactions on Learning Technologies, 9*(3), 217-230. https://doi.org/10.1109/TLT.2016.2521649

*Towers, A., Field, J., Stokes, C., Maddock, S., & Martin, N. (2019). A scoping review of the use and application of virtual reality in pre-clinical dental education. *British dental journal, 226*(5), 358-366. https://doi.org/10.1038/s41415-019-0041-0

*Troussas, C., Krouska, A., & Sgouropoulou, C. (2020). Collaboration and fuzzy-modeled personalization for mobile game-based learning in higher education. *Computers & Education, 144*, 103698, 1-18. https://doi.org/10.1016/j.compedu.2019.103698

Turing, A. (1950). Computing machinery and intelligence. *Mind: A quarterly review of psychology and philosophy. 59*(236), 433-460. https://doi.org/10.1093/mind/LIX.236.433

*Van der Meij, H., Van der Meij, J., & Harmsen, R. (2015). Animated pedagogical agents effects on enhancing student motivation and learning in a science inquiry environment. *Educational*

Technology Research and Development, 63, 381–403. http://dx.doi.org/10.1007/s11423-015-9378-5

*Veletsianos, G., & Russell, G. S. (2013). What do learners and pedagogical agents discuss when given opportunities for open-ended dialogue?. *Journal of Educational Computing Research, 48*(3), 381-401. https://doi.org/10.2190/EC.48.3.e

Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes*. Harvard university press.

Wallace, R.S. (2009). The Anatomy of A.L.I.C.E.. In Epstein, R., Roberts, G., Beber, G. (Eds). *Parsing the Turing Test* (pp. 181-210). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-6710-5_13

Wang H. & Zheng H. (2013). Model Validation, Machine Learning. In Dubitzky W., Wolkenhauer O., Cho KH., Yokota H. (Eds). *Encyclopedia of Systems Biology* (pp. 1406-1407). Springer, New York, NY. https://doi.org/10.1007/978-1-4419-9863-7_233

*Westera, W., Dascalu, M., Kurvers, H., Ruseti, S., & Trausan-Matu, S. (2018). Automated essay scoring in applied games: Reducing the teacher bandwidth problem in online training. *Computers & Education, 123*, 212-224. https://doi.org/10.1016/j.compedu.2018.05.010

*Winkler-Schwartz, A., Bissonnette, V., Mirchi, N., Ponnudurai, N., Yilmaz, R., Ledwos, N., ... & Del Maestro, R. F. (2019). Artificial intelligence in medical education: best practices using machine learning to assess surgical expertise in virtual reality simulation. *Journal of surgical education, 76*(6), 1681-1690. https://doi.org/10.1016/j.jsurg.2019.05.015

Wohlin, C., Runeson, P., da Mota Silveira Neto, P. A., Engström, E., do Carmo Machado, I., & Santana de Almeida, E. (2013). On the Reliability of Mapping Studies in Software

Engineering. *Journal of Systems and Software, 86*(10), 2594-2610. https://doi.org/10.1016/j.jss.2013.04.076

Woolf, B. P., Lane, H. C., Chaudhri, V. K., & Kolodner, J. L. (2013). AI grand challenges for education. *AI magazine, 34*(4), 66-84. https://doi.org/10.1609/aimag.v34i4.2490

Xu, J., Szlam, A., & Weston, J. (2021). Beyond goldfish memory: Long-term open-domain conversation. *arXiv preprint*. https://arxiv.org/abs/2107.07567

*Yang, J., Thomas, M. S., Qi, X., & Liu, X. (2019). Using an ANN-based computational model to simulate and evaluate Chinese students' individualized cognitive abilities important in their English acquisition. *Computer Assisted Language Learning, 32*(4), 366-397. https://doi.org/10.1080/09588221.2018.1517125

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, 100025, 1-11. https://doi.org/10.1016/j.caeai.2021.100025

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational psychologist, 25*(1), 3-17. https://doi.org/10.1207/s15326985ep2501_2

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: