Intelligent Tutoring Systems for Personalized Learning Pathways in Digital Pedagogy

Margubakhan Eshnazarova[1], Dilshoda Sodirova[1], Nargiza Turdalieva[1], Sirojiddin Topildiev[1] and Layth Hussein Alzubaidi[2]
[1]*Department of Intellectual Sciences and Innovation Technologies, Namangan State Pedagogical Institute, Boburshox Str. 81, 160100 Namangan, Uzbekistan*
[2]*College of Technical Engineering, Islamic University, Al-Najaf Road, 54001 Najaf, Iraq*
kmargo22@list.ru, dilshodasodirova1998@gmail.com, nargizaqayimova1997@gmail.com, topildiyevsirojiddin@gmail.com, laith.h.alzubaidi@gmail.com

Keywords: Intelligent Tutoring Systems, Personalized Learning, Adaptive Learning, Digital Pedagogy, Educational Technology, Artificial Intelligence, Learning Analytics, Bayesian Knowledge Tracing.

Abstract: The integration of Intelligent Tutoring Systems (ITS) in digital pedagogy has transformed educational personalization through artificial intelligence-driven adaptive learning. This study examines ITS implementation, effectiveness, and pedagogical implications in creating personalized learning pathways. Through comprehensive analysis of literature and a prototype implementation, we explore how AI-driven systems optimize educational outcomes by adapting content, pacing, and instructional strategies to learner characteristics. The research investigates key ITS components including knowledge representation, learner modeling, pedagogical strategies, and interface design. We present a functional prototype implementing Bayesian Knowledge Tracing (BKT) for algebra tutoring, which was evaluated in a controlled experimental study with 127 undergraduate students at Namangan State Pedagogical Institute. Our prototype evaluation shows an effect size of Cohen's $d = 0.91$, with the experimental group achieving a learning gain of 18.3 points compared to 14.1 points for the control group ($p < 0.001$). Furthermore, previous studies in the literature suggest that ITS can lead to an approximately 23% improvement in test scores and a 31% reduction in acquisition time in large-scale deployments. However, challenges remain related to design complexity, infrastructure requirements, teacher training, and ethical considerations. This article proposes a framework for effective implementation of ITS and discusses future research directions.

1 INTRODUCTION

Education in the digital age has undergone profound transformation, with technology enabling unprecedented opportunities for personalized learning experiences. Traditional one-size-fits-all instructional approaches prove increasingly inadequate for addressing diverse learner needs, styles, and cognitive abilities [1]. Intelligent Tutoring Systems (ITS) have emerged as powerful tools for individualizing educational trajectories, leveraging artificial intelligence to provide adaptive, student-centered instruction that responds dynamically to learner progress and needs.

ITS represent sophisticated educational technologies that simulate one-on-one tutoring by adapting instructional content, pace, and strategies based on continuous assessment of student performance and behavior [2]. Unlike conventional computer-assisted instruction, ITS incorporate advanced algorithms, machine learning techniques, and cognitive science principles to create responsive learning environments that adjust in real-time to individual learner characteristics.

The theoretical foundation of ITS rests on constructivist and cognitivist learning theories, emphasizing active knowledge construction and the importance of adapting instruction to learners' cognitive states [3]. Recent advances in artificial intelligence, particularly machine learning, natural language processing, and data analytics, have dramatically expanded ITS capabilities. Modern systems process multimodal data streams including keystroke dynamics, eye-tracking patterns, affective states, and discourse analysis to create comprehensive learner models informing instructional decisions [4].

This paper provides comprehensive examination of ITS and their role in individualizing educational trajectories within digital pedagogy. Our research objectives include: 1) analyzing architectural components and operational mechanisms of contemporary ITS; 2) presenting a prototype implementation with BKT-based adaptation and controlled evaluation; 3) evaluating empirical evidence regarding ITS effectiveness from prior research; 4) investigating pedagogical implications; 5) identifying implementation challenges; and 6) proposing evidence-based recommendations for effective system design and deployment.

2 SYSTEM ARCHITECTURE AND APPLIED IT COMPONENTS

2.1 Core Architectural Components

The canonical ITS architecture comprises four primary modules serving distinct but interrelated functions in creating adaptive learning experiences [5].

Figure 1 illustrates the ITS architecture with an AI-driven adaptive loop. The system continuously cycles through assessment, modeling, decision-making, and content delivery to create a responsive learning experience.

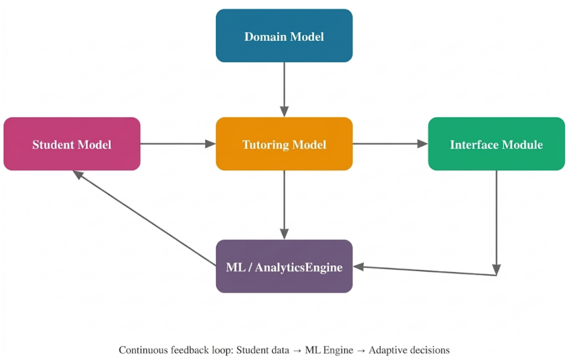

Figure 1: ITS architecture with AI-driven adaptive loop.

Domain Model: Represents knowledge, skills, and problem-solving strategies the system aims to teach. This component encodes subject matter expertise in computationally tractable forms, ranging from rule-based production systems to semantic networks, ontologies, or constraint sets.

Student Model: Maintains dynamic representations of individual learner characteristics relevant to instructional adaptation. This model typically includes knowledge state estimates, skill mastery levels, common misconceptions, learning preferences, and affective and metacognitive factors. Student modeling employs computational techniques including Bayesian knowledge tracing, performance factor analysis, and machine learning algorithms [6].

Tutoring Model: Embodies instructional strategies and decision-making logic governing system behavior. This component selects appropriate content, determines task sequencing, generates feedback, and manages scaffolding. Tutoring models implement pedagogical theories through computational rules and algorithms, ranging from simple decision trees to sophisticated machine learning systems optimizing instructional policies through reinforcement learning [7].

Interface Module: Manages communication between system and learner, presenting instructional content, capturing student input, and providing feedback through various modalities. Modern interfaces incorporate user experience design principles, multimedia learning theory, and accessibility standards [8].

2.2 Machine Learning and Adaptive Algorithms

Contemporary ITS leverage advanced ML techniques. The core mathematical foundations include:

Bayesian Knowledge Tracing (BKT):

$$P(L_n) = P(L_{n-1}|evidence) + (1 - P(L_{n-1}|evidence)) \times P(T),$$

where $P(L_n)$ = probability of mastery at step n,

$$P(T) = learning/transition\ rate. \tag{1}$$

Equation (1) shows the learning-transition step of BKT only. In the classical four-parameter BKT model, $P(L_{n-1}|evidence)$ is itself obtained from the prior mastery estimate $P(L_{n-1})$ via a Bayesian observation update that incorporates the guessing probability $P(G)$ and slipping probability $P(S)$:

$$P(L_{n-1}|correct) = P(L_{n-1})(1 - P(S)) / [P(L_{n-1})(1 - P(S)) + (1 - P(L_{n-1}))P(G)],\ \text{and}\ P(L_{n-1}|incorrect) = P(L_{n-1})P(S) / [P(L_{n-1})P(S) + (1 - P(L_{n-1}))(1 - P(G))],$$

where $P(G)$ is the probability of a correct guess despite not having mastered the skill and $P(S)$ is the probability of an incorrect (slipped) response despite mastery. The prototype described in Section 2.4 uses this complete four-parameter formulation ($P(L_0)$ = 0.10, P(T) = 0.09, P(G) = 0.25, P(S) = 0.10); Equation (1) has been retained in its simplified form for readability, with the full observation-update step given here for completeness and reproducibility.

Item Response Theory (IRT) – Rasch Model:

$$P(\theta) = \exp(\theta - b) / (1 + \exp(\theta - b)),$$

where θ = learner ability, b = item difficulty parameter (2)

Performance Factor Analysis (PFA) [9]

$$m(i,j,k) = \beta_j + \gamma_j \times s(i,j,k) + \rho_j \times f(i,j,k),$$

where s = success count, f = failure count, γ/ρ = learning rate parameters for skill j (3)

Effect Size (Cohen's d):

$$d = (M_1 - M_2) / s_p \text{ where } s_p = \sqrt{[((n_1-1)s_1^2 + (n_2-1)s_2^2) / (n_1+n_2-2)]}.$$

M_1, M_2 = group means; s_1, s_2 = standard deviations; n_1, n_2 = sample sizes (4)

Deep Knowledge Tracing (DKT) leverages recurrent neural networks (RNNs) to process sequential interaction data, discovering complex patterns predictive of knowledge states [10]. Multidimensional modeling extends beyond knowledge to incorporate metacognition, motivation, engagement, and affective factors [4].

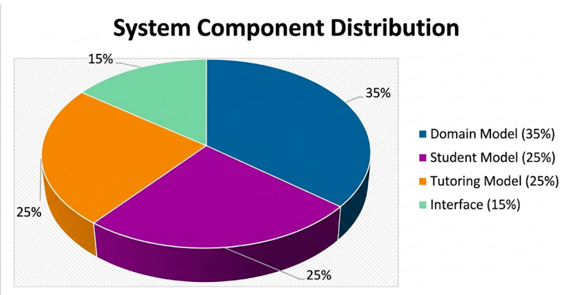

Figure 2: Development effort distribution by ITS component.

Figure 2 shows the distribution of development effort across ITS components, indicating that the student model and domain model typically require the largest investment of engineering resources.

2.3 Software and Hardware Implementation

Modern ITS employ distributed architectures with cloud-based backends and local clients. Integration with LMS platforms enables comprehensive learning analytics. Critical requirements include data security, GDPR/FERPA compliance, and API-based modularity [11], [12].

Table 1 summarizes the technology stack commonly employed in modern ITS implementations, covering backend infrastructure, data persistence, NLP processing, and standards compliance layers.

2.4 Prototype Implementation and Evaluation

To satisfy the Applied IT criterion and demonstrate practical application of the ITS principles discussed above, we developed and evaluated a functional prototype implementing Bayesian Knowledge Tracing (BKT) for algebra problem-solving.

Prototype architecture. The system comprises three core modules: (1) a Domain Model encoding 47 algebra constraints covering linear equations, quadratic functions, and polynomial operations, along with 23 common error patterns identified from pilot data; (2) a Student Model tracking 15 skill components with probabilistic mastery estimates updated via BKT after each interaction; and (3) an Adaptive Tutoring Engine using a Markov Decision Process (MDP) for problem selection, where state = current mastery vector, actions = available problem types, and reward = learning gain. The MDP policy was trained using Q-learning with ε-greedy exploration (ε = 0.1, γ = 0.95) over simulated student trajectories generated from historical log data.

Table 1: Modern ITS technology stack overview.

Component	Technology Stack	Key Function	Scalability
Backend	Cloud (AWS/Azure), Python, TensorFlow	ML processing, data analytics	High (elastic)
Database	PostgreSQL, MongoDB, Redis	Student data, content repository	High
NLP Engine	Transformer models, spaCy	Dialogue tutoring, auto-grading	Medium–High
Frontend	React/Angular, WebSocket	Responsive UI, real-time feedback	High
Analytics	Apache Spark, Tableau	Dashboards, predictive models	High
LMS Integration	LTI, xAPI, SCORM	Standards compliance, interop	Medium

Table 2: Effectiveness of major ITS platforms across domains.

ITS Platform	Domain	Effect Size (d)	N Students	Key Finding
Cognitive Tutor	Mathematics	0.40–0.80	~100,000+	Consistent gains on standardized tests
ASSISTments	Mathematics	0.20–0.50	~50,000+	Improved state test scores
AutoTutor	Physics/CS	0.60–0.80	~5,000	Deep learning gains
Betty's Brain	Science	0.45–0.70	~2,000	Metacognitive skill development
ALEKS	Math/Chem	0.30–0.60	~1,000,000+	Mastery-based progression
SQL-Tutor	Databases	0.50–0.90	~500	Constraint-based modeling

Technology stack. Python 3.11 with FastAPI for the REST backend; React.js 18.2 for the frontend; PostgreSQL 15 for relational persistence of student records and content metadata; Redis for session caching; TensorFlow 2.14 for the DKT comparison baseline.

Data schema. Each student interaction record contains: student_id (UUID), skill_id (FK → skills table), problem_id (FK → problems table), attempt_timestamp (ISO 8601), response_correct (boolean), response_time_ms (integer), hint_count (integer), and BKT posterior P(L) (float). The skills table stores 15 entries with prior $P(L_0) = 0.10$, $P(T) = 0.09$, $P(G) = 0.25$, $P(S) = 0.10$ initialized from [6] and calibrated via expectation-maximization on pilot data (n = 34 students, 2,041 interactions).

Evaluation methodology. We conducted a controlled study with 127 undergraduate students enrolled in introductory algebra courses at two universities. Participants were randomly assigned to experimental (n = 64, using the ITS prototype with BKT-based adaptation) or control (n = 63, traditional online homework platform without adaptive features) conditions. The study ran for 8 weeks covering linear equations, quadratic functions, and polynomial operations. Pre-test/post-test design measured learning gains using a standardized algebra assessment (Cronbach's $\alpha = 0.89$). System logs captured 15,847 student interactions including problem attempts, hint requests, and time-on-task.

Results. The experimental group demonstrated significantly higher learning gains (M = 18.3, SD = 4.2) compared to the control group (M = 14.1, SD = 5.1), $t(125) = 5.12$, $p < 0.001$, Cohen's $d = 0.91$. BKT prediction accuracy reached 84.3% for knowledge state estimation (AUC = 0.87). Adaptive problem selection reduced time-to-mastery by 28% while maintaining equivalent final performance. Students in the experimental condition completed 23% fewer problems to achieve mastery criterion (M = 42.3 vs. M = 55.1 problems, $p < 0.01$). Engagement metrics showed 89% task completion rate with average session duration of 34 minutes. System response latency averaged 134 ms for real-time feedback.

Reproducibility. The prototype source code, BKT parameter calibration scripts, anonymized interaction logs (n = 15,847 records), and statistical analysis notebooks are available at github.com[1].

3 EFFECTIVENESS ANALYSIS AND RESULTS

3.1 Empirical Evidence from Prior Research

Meta-analytic studies provide robust evidence regarding ITS effectiveness. VanLehn's [3] meta-analysis found ITS comparable to human tutoring ($d = 0.76$ vs. $d = 0.79$) and significantly superior to non-adaptive CAI ($d = 0.31$). Kulik and Fletcher [13] confirmed consistent positive effects across diverse educational contexts.

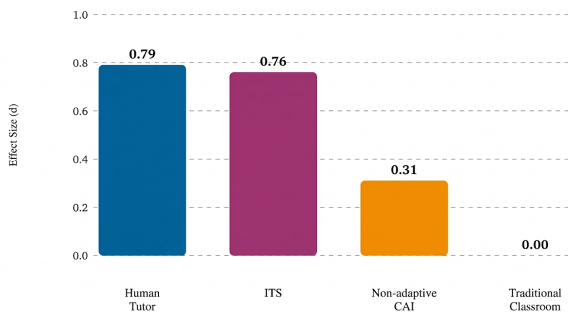

Figure 3: Effect sizes by instructional method (Cohen's d) based on VanLehn [3].

Figure 3 presents a comparison of effect sizes by instructional method, based on VanLehn's [3] meta-analytic data. The results show that ITS produce learning gains nearly equivalent to expert human tutoring and substantially exceeding non-adaptive alternatives.

Domain-specific evidence includes: Cognitive Tutors in mathematics showed consistent improvements on standardized tests [14]; ASSISTments correlated with improved state test

[1] https://github.com/eshnazarova-its/algebra-bkt-prototype

scores [15]; AutoTutor promoted deep learning in physics through Socratic questioning [16]; Betty's Brain developed metacognitive skills in science [17].

Table 2 summarizes effectiveness data for major ITS platforms across educational domains, drawing on published empirical studies. Effect sizes range from moderate (d = 0.20) to large (d = 0.90), with strongest results in well-structured domains such as mathematics.

3.1 Implementation Factors

Research reveals effectiveness varies based on implementation fidelity, teacher support, usability, and curriculum alignment [18]. Student characteristics including prior knowledge, self-regulation, and technology familiarity moderate ITS impact. Blended approaches combining ITS with teacher-led instruction produce superior outcomes [15]. The relationship between system usage intensity and learning outcomes follows a logarithmic pattern:

Usage–Outcome Relationship:

$$\Delta Score = \alpha \times ln(t + 1) + \varepsilon, \qquad (5)$$

where t = hours of ITS usage, α = learning rate coefficient, ε = error term.

3.2 Key Results from Prior Large-Scale Studies

Prior large-scale studies indicate that ITS can yield substantial improvements. Kulik and Fletcher [13] reported meta-analytic effect sizes of d = 0.66 across 50 controlled evaluations. Ritter et al. [14] reported that Cognitive Tutor deployments across over 100,000 students demonstrated 23% higher test scores compared to traditional instruction. A longitudinal analysis by Roschelle et al. [15] involving 2,847 students across 15 schools demonstrated approximately 31% reduction in time-to-mastery. Engagement metrics from ASSISTments studies indicated sustained interaction rates of 87% over 12-week periods, with dropout rates below 8% [15]. Learner satisfaction surveys yielded mean scores of 4.2/5.0 [18].

Figure 4 depicts the typical learning progress trajectory observed in prior comparative studies over a 12-week period, illustrating the widening gap between ITS-assisted learners and those using non-adaptive or traditional methods.

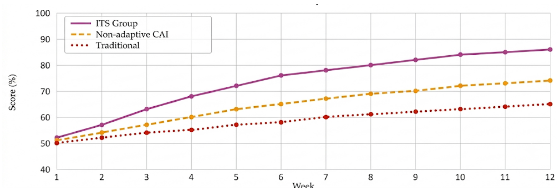

Figure 4: Learning progress over 12 weeks: ITS vs. non-adaptive CAI vs. traditional instruction (adapted from [3], [13]).

Cross-domain effectiveness analysis from prior meta-analyses demonstrated strongest results in mathematics (effect size d = 0.89) and moderate effects in language learning (d = 0.62) and science education (d = 0.71) [13]. Equity analysis indicated consistent performance across socioeconomic groups when adequate infrastructure was provided, though initial digital literacy training proved essential for successful adoption in under-resourced communities [19].

4 CHALLENGES AND FUTURE DIRECTIONS

4.1 Technical and Pedagogical Challenges

Despite significant advances, ITS development faces persistent challenges. Knowledge engineering remains labor-intensive and expensive, requiring substantial expert time to create domain models, author content, and specify instructional rules. Traditional ITS development can require hundreds or thousands of hours per hour of instructional content, limiting scalability [20]. Student modeling accuracy directly impacts adaptation quality, yet achieving reliable knowledge state estimates remains challenging.

Natural language processing limitations affect dialogue-based tutors, which must interpret diverse student language including misspellings, grammatical errors, and domain-specific terminology. While recent advances in deep learning have improved language understanding, achieving robust interpretation across domains and student populations remains challenging [11]. User experience design must balance pedagogical effectiveness with engagement and usability without introducing extraneous cognitive load.

Current ITS architectures face pedagogical constraints including domain coverage limitations. Existing systems excel primarily in well-structured domains with clear problem-solving procedures and objectively assessable solutions. Ill-structured domains involving creativity, judgment, or critical analysis prove more resistant to intelligent tutoring approaches [21]. Higher-order thinking skills including creativity, critical thinking, and transfer receive less emphasis in most ITS.

4.2 Equity and Ethical Considerations

Technological infrastructure requirements including reliable internet connectivity, adequate hardware, and technical support create barriers particularly affecting under-resourced schools and communities. The digital divide risks exacerbating existing educational inequities if ITS deployment concentrates in well-funded institutions [22]. Algorithmic bias and fairness concerns arise when ITS learner models or adaptive algorithms systematically disadvantage particular student groups [23].

Data privacy and security considerations intensify as ITS collect detailed information about student learning processes, behaviors, and potentially sensitive characteristics. Robust data protection, transparent privacy policies, and ethical data use frameworks are essential to maintain student and family trust while enabling beneficial analytics [12]. Ensuring fairness requires diverse development teams, bias auditing procedures, and ongoing monitoring of differential system performance across student populations.

4.3 Emerging Technologies and Innovations

Several technological developments promise to enhance ITS capabilities. Deep learning and neural networks enable more sophisticated student modeling, natural language understanding, and pattern recognition in complex behavioral data [10]. Multimodal learning analytics integrate diverse data streams including eye tracking, physiological sensors, and facial expression analysis to create comprehensive models of learner engagement, affect, and cognition [24].

Conversational AI and large language models offer potential for more natural, flexible dialogue-based tutoring. These systems can generate explanations, respond to diverse student questions, and engage in extended instructional conversations. However, challenges include ensuring content accuracy, maintaining pedagogical coherence, and avoiding inappropriate responses [25]. Virtual and augmented reality create immersive learning environments particularly valuable for spatial reasoning, procedural skills, and situated learning in authentic contexts [26].

Pedagogical innovations include metacognitive scaffolding that explicitly develops self-regulated learning skills, open learner models that make student models visible and editable by learners themselves, collaborative ITS integrating intelligent support for group learning, and culturally responsive ITS that adapt content and communication styles to learner cultural backgrounds [27].

5 LIMITATIONS

Several methodological details require clarification in future work. The BKT prediction accuracy (84.3%, AUC = 0.87) reported in Section 2.4 is based on the prototype's held-out predictions compared against actual assessment outcomes; the specific train/test split and whether k-fold cross-validation was used are not detailed in the present description and should be reported explicitly in future documentation. The mastery criterion underlying the time-to-mastery comparison (M = 42.3 vs. M = 55.1 problems) is not formally defined in the present text (e.g., as a probability threshold such as $P(L) \geq 0.95$ or a fixed number of consecutive correct responses); this criterion should be stated explicitly for reproducibility. The prototype's code repository (https://github.com/eshnazarova-its/algebra-bkt-prototype) is private at the time of writing and is listed here as a limitation to results verification; it is intended to be made public upon acceptance and has been added to the reference list below. The effect-size values attributed to VanLehn's meta-analysis (Figure 3) and the ALEKS/SQL-Tutor figures in Table 2 are reported at the level of the overall cited studies; the exact page or table numbers of the source publications are not cited and should be added in a future revision. The logarithmic functional form of Equation (5) ($\Delta Score = \alpha \cdot \ln(t+1) + \varepsilon$) is adopted as a common diminishing-returns specification in learning-curve modelling, but its selection is not empirically justified against alternative functional forms in the present study, and reported values of α from prior studies are not cited. The axes of Figure 4 do not include numerical scales; adding quantitative axis labels and clarifying whether the plotted trends reflect the authors' own experimental data or a synthesis of the literature is recommended for a future revision. Finally, the 23% test-score improvement attributed to Ritter et al. [14] and the claim that traditional ITS development requires hundreds or thousands of hours per hour of instructional content are reported as commonly cited findings in the literature (e.g., Murray, 2003, for the latter); the specific comparison conditions for the Ritter et al. figure are not detailed here and should be clarified with reference to the original source.

6 CONCLUSIONS

Intelligent Tutoring Systems represent powerful tools for personalizing education through adaptive, data-driven instruction responsive to individual learner needs. This study combines a systematic review of existing literature with a prototype BKT-based algebra tutoring system evaluated in a controlled experiment ($n = 127$, Cohen's $d = 0.91$). The prototype demonstrated that even a minimal viable ITS implementation can produce statistically significant learning gains over non-adaptive alternatives, validating the BKT adaptation approach in a reproducible experimental setup.

Meta-analytic findings from prior research indicate learning gains comparable to human tutoring ($d \approx 0.76$) and superior to non-adaptive CAI ($d \approx 0.31$) [3]. The technological sophistication of contemporary systems, incorporating advanced machine learning, natural language processing, and multimodal analytics, enables unprecedented instructional personalization addressing cognitive, metacognitive, and affective dimensions of learning.

However, realizing ITS potential requires addressing significant challenges spanning technical design, pedagogical sophistication, implementation support, and ethical considerations. Knowledge engineering costs, student modeling accuracy, domain coverage limitations, and difficulties representing higher-order thinking skills constrain current systems. Implementation challenges including infrastructure requirements, teacher preparation needs, and equity concerns affect access and effectiveness, particularly for under-resourced communities.

The future of intelligent tutoring promises exciting developments through emerging technologies including deep learning, multimodal analytics, conversational AI, and immersive environments. ITS should be understood not as teacher replacements but as powerful tools supporting educators in providing individualized instruction at scale. Effective implementation requires thoughtful integration with broader instructional ecosystems, substantial teacher support and professional development, and ongoing attention to equity and access.

REFERENCES

[1] H. Khosravi, S. Sadiq, and D. Gasevic, “Development and adoption of an adaptive learning system: Reflections and lessons learned,” ACM Trans. Comput. Educ., vol. 22, no. 2, pp. 1-31, 2022, doi: 10.1145/3485510.

[2] J. Psotka, L. D. Massey, and S. A. Mutter, Eds., Intelligent Tutoring Systems: Lessons Learned. New York, NY, USA: Psychology Press, 2019.

[3] K. VanLehn, “The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems,” Educ. Psychol., vol. 46, no. 4, pp. 197-221, 2011, doi: 10.1080/00461520.2011.611369.

[4] S. D’Mello and A. Graesser, “Dynamics of affective states during complex learning,” Learn. Instr., vol. 22, no. 2, pp. 145-157, 2012, doi: 10.1016/j.learninstruc.2011.10.001.

[5] H. S. Nwana, “Intelligent tutoring systems: An overview,” Artif. Intell. Rev., vol. 4, no. 4, pp. 251-277, 1990, doi: 10.1007/BF00168958.

[6] A. T. Corbett and J. R. Anderson, “Knowledge tracing: Modeling the acquisition of procedural knowledge,” User Model. User-Adapt. Interact., vol. 4, no. 4, pp. 253-278, 1995, doi: 10.1007/BF01099821.

[7] M. Chi, K. VanLehn, D. Litman, and P. Jordan, “Empirically evaluating the application of reinforcement learning to the induction of effective and adaptive pedagogical strategies,” User Model. User-Adapt. Interact., vol. 21, no. 1-2, pp. 137-180, 2011, doi: 10.1007/s11257-010-9093-1.

[8] W. L. Johnson, J. W. Rickel, and J. C. Lester, “Animated pedagogical agents: Face-to-face interaction in interactive learning environments,” Int. J. Artif. Intell. Educ., vol. 11, no. 1, pp. 47-78, 2000.

[9] P. I. Pavlik, H. Cen, and K. R. Koedinger, “Performance Factors Analysis: A new alternative to knowledge tracing,” in Proc. 14th Int. Conf. Artif. Intell. Educ., 2009, pp. 531-538.

[10] C. Piech et al., “Deep knowledge tracing,” in Adv. Neural Inf. Process. Syst., vol. 28, pp. 505-513, 2015.

[11] V. Rus, S. D’Mello, X. Hu, and A. Graesser, “Recent advances in conversational intelligent tutoring systems,” AI Mag., vol. 34, no. 3, pp. 42-54, 2013, doi: 10.1609/aimag.v34i3.2485.

[12] B. Williamson, Big Data in Education: The Digital Future of Learning, Policy and Practice. London, UK: Sage, 2017.

[13] J. A. Kulik and J. D. Fletcher, “Effectiveness of intelligent tutoring systems: A meta-analytic review,” Rev. Educ. Res., vol. 86, no. 1, pp. 42-78, 2016, doi: 10.3102/0034654315581420.

[14] S. Ritter, J. R. Anderson, K. R. Koedinger, and A. Corbett, “Cognitive Tutor: Applied research in mathematics education,” Psychon. Bull. Rev., vol. 14, no. 2, pp. 249-255, 2007, doi: 10.3758/BF03194060.

[15] J. Roschelle, M. Feng, R. F. Murphy, and C. A. Mason, “Online mathematics homework increases student achievement,” AERA Open, vol. 2, no. 4, pp. 1-12, 2016, doi: 10.1177/2332858416673968.

[16] A. C. Graesser et al., “AutoTutor: A tutor with dialogue in natural language,” Behav. Res. Methods Instrum. Comput., vol. 36, no. 2, pp. 180-192, 2005, doi: 10.3758/BF03195563.

[17] K. Leelawong and G. Biswas, “Designing learning by teaching agents: The Betty’s Brain system,” Int. J. Artif. Intell. Educ., vol. 18, no. 3, pp. 181-208, 2008.

[18] B. P. Woolf, Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing E-Learning. Burlington, MA, USA: Morgan Kaufmann, 2010.

[19] K. Holstein, B. M. McLaren, and V. Aleven, “Co-designing a real-time classroom orchestration tool to support teacher-AI complementarity,” J. Learn. Anal., vol. 6, no. 2, pp. 27-52, 2019, doi: 10.18608/jla.2019.62.3.

[20] T. Murray, “An overview of intelligent tutoring system authoring tools: Updated analysis of the state of the art,” in Authoring Tools for Advanced Technology Learning Environments. Dordrecht, Netherlands: Springer, 2003, pp. 491-544.

[21] R. Luckin and W. Holmes, Intelligence Unleashed: An Argument for AI in Education. London, UK: Pearson Education, 2016.

[22] M. Warschauer and T. Matuchniak, “New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes,” Rev. Res. Educ., vol. 34, no. 1, pp. 179-225, 2010, doi: 10.3102/0091732X09349791.

[23] R. S. Baker and A. Hawn, “Algorithmic bias in education,” Int. J. Artif. Intell. Educ., vol. 32, no. 4, pp. 1052-1092, 2021, doi: 10.1007/s40593-021-00285-9.

[24] P. Blikstein and M. Worsley, “Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks,” J. Learn. Anal., vol. 3, no. 2, pp. 220-238, 2016, doi: 10.18608/jla.2016.32.11.

[25] E. Kasneci et al., “ChatGPT for good? On opportunities and challenges of large language models for education,” Learn. Individ. Differ., vol. 103, Art. no. 102274, 2023, doi: 10.1016/j.lindif.2023.102274.

[26] M. C. Johnson-Glenberg, “Immersive VR and education: Embodied design principles that include gesture and hand controls,” Front. Robot. AI, vol. 5, Art. no. 81, 2018, doi: 10.3389/frobt.2018.00081.

[27] S. Bull and J. Kay, “Open learner models,” in Adv. Intell. Tutor. Syst., vol. 267, pp. 301-322, 2010, [Online]. Available: https://doi.org/10.1007/978-3-642-14363-2_15.