The Algorithm Knows Best? Challenges and Advances in Self-Paced Adaptive Learning for K–12

Elisa de Oliveira Flemer[1] and Flavia Maria Santoro[1]

Inteli – Instituto de Tecnologia e Liderança, São Paulo, Brazil
{elisa.flemersou.inteli.edu.br, flaviainteli.edu.br

Abstract. This systematic review investigates recent advances in AI-driven, self-paced adaptive learning systems tailored for K–12 education. While adaptive technologies have gained traction in higher education, K–12 contexts remain underexplored despite their unique pedagogical needs and potential for impact. From an initial pool of 1,031 articles, we analyzed 31 studies published since 2020 that met strict inclusion criteria—including algorithm-driven personalization, core academic focus, and empirical validation.
We categorize methods across student modeling, domain knowledge representation, personalization mechanisms, and evaluation strategies. Our findings reveal a growing shift toward deep learning and hybrid architectures, which outperform classical models in prediction accuracy but often sacrifice interpretability. Manual labeling remains prevalent, despite advances in contrastive learning and semantic modeling. Learning styles, while still used, lack empirical grounding. Critically, few systems undergo robust real-world experimentation—only two employed randomized controlled trials.
Based on this synthesis, we highlight urgent directions for research: developing explainable models, automating curriculum-aligned knowledge extraction, integrating LLMs responsibly for content generation, and conducting rigorous in-situ validation. By advancing these areas, adaptive learning systems can move beyond performance prediction to foster transparency, motivation, and genuine pedagogical alignment in K–12 settings

Keywords: Adaptive education. · K-12. · Artificial intelligence.

1 Introduction

In an adaptive learning environment, a student begins the school day with material calibrated to their current level of understanding—advancing at their own pace without being held back or left behind. Instruction is personalized and responsive, offering targeted feedback and support as needed. Meanwhile, the teacher is no longer confined to routine correction or uniform pacing, and instead can focus on mentoring, relationship-building, and timely intervention—guided by real-time insights into student progress. This approach, while increasingly feasible through AI, remains underutilized and insufficiently studied in K–12 education.

Adaptive learning has long fascinated educators, tracing back to B. F. Skinner's conceptual teaching machines [36] and now reaching new heights with the revolution of Large Language Models (LLMs). In this context, the personalization of education has emerged as a pivotal objective in contemporary educational practices, particularly with AI-driven adaptive learning systems, which consistently demonstrate better results in student engagement and retention compared to traditional teacher-led methods [19]. However, most existing research on such systems, especially with recent advances in Generative AI, has predominantly focused on higher education contexts, where students typically exhibit greater independence and intrinsic motivation. Consequently, K-12 educational environments remain relatively underexplored, often lacking rigorous empirical validation [**?**].

This gap presents significant challenges for leveraging personalized learning at a crucial developmental stage. After all, personalized learning strategies have been identified as potential solutions to high dropout rates, which are notoriously influenced by a lack of student motivation and engagement [8]. As such, understanding how adaptive learning practices can specifically benefit the unique profile of K-12 students—as opposed to higher education—is critical. Hence, addressing this issue requires systematically synthesizing existing research to identify effective adaptive learning practices and computational methods tailored specifically to K-12 environments.

In this sense, this systematic review seeks to address the following research questions:

– What student modeling techniques, domain knowledge representations, and personalization mechanisms are currently used in AI-driven adaptive learning systems for K–12 education?
– How have these systems been implemented and evaluated in practice, and what outcomes have been reported regarding learning performance, engagement, and system robustness in K–12 contexts?
– What open challenges and future research directions emerge from recent studies on AI-driven adaptive learning in K–12 education, particularly regarding validation methods, scalability, and alignment with pedagogical needs?

By addressing these questions, this systematic review provides a foundational synthesis of current research on adaptive education in K-12 settings. Due to its structured approach, this paper ensures transparency, reproducibility, and reduced bias—advantages less achievable through traditional narrative. As a result, the insights derived from this analysis will guide educators, researchers, and policymakers in developing practical, effective strategies for enhancing personalized learning in K-12 education.

The rest of the paper is structured as follows: Section 2 outlines the systematic review methodology, including the search strategy, inclusion and exclusion criteria, and screening process. Section 3 presents the results, beginning with study selection and an overview of the selected works, followed by a synthesis of student modeling approaches, knowledge representation frameworks, and personalization mechanisms. Section 4 discusses implementation and evaluation practices, including subject domains, empirical validation methods, and performance outcomes. Section 5 offers a critical discussion of key challenges such as manual labeling, explainability, motivation modeling, and the integration of deep learning and large language models. Finally, Section 6 concludes with reflections on future research directions and the broader implications for K–12 adaptive learning.

2 Methods

2.1 Search Strategy

This systematic review was conducted using a rigorous search protocol to identify and analyze recent studies on AI-based adaptive learning in K-12 education. Our primary focus was on systems that autonomously teach students from their first contact with new content, rather than serving as supplemental tools or teacher aids.

To ensure comprehensive coverage of relevant literature, we formulated the following search string:

> ("adaptive learning" OR "personalized learning") AND ("K-12" OR "elementary school" OR "middle school" OR "high school") AND ("machine learning" OR "artificial intelligence")

This search string was executed in the ACM Guide to Computing Literature and the IEEE Xplore Digital Library, both filtered for studies published from 2020 onwards. These databases were selected due to their extensive coverage of computer science and AI research. The terms were searched across all default metadata fields in each respective search engine.

2.2 Inclusion and Exclusion Criteria

To ensure relevance and methodological rigor, we applied a two-stage filtering process: initial screening (title and abstract) and full-text analysis. As our understanding of the research landscape evolved, we refined our criteria to focus more precisely on studies that aligned with our objective of self-paced adaptive learning.

Initially, we employed broader inclusion criteria, selecting studies that contained empirical validation, focused on K-12 education, and were published from 2020 onwards. We also prioritized studies that examined primary teaching rather than after-hours, supplemental, or tutoring-based approaches.

However, as we examined the literature, we found a vast number of niche subfields that did not align with our goal of fully AI-driven, self-paced adaptive learning. Many studies focused on wearable devices, classroom-based interventions, or demographic-based adaptation rather than direct AI-powered learning on standard student devices such as computers, tablets, or mobile phones.

Furthermore, we observed that AI-based adaptive learning encompassed multiple subdomains, such as student modeling, knowledge representation, knowledge tracing, recommendation systems, and automatic question generation. To refine our focus, we adjusted our criteria to capture all critical aspects of self-paced adaptive learning while excluding research reliant on external gadgets, teacher-centric approaches, or demographic-based adaptation.

Inclusion Criteria Selected studies had to meet all the following criteria:

- *Use of AI/ML:* The study must explicitly incorporate artificial intelligence or machine learning techniques in adaptive learning, ensuring an algorithm-driven approach to content personalization.
- *Core K-12 Subjects:* The study must focus on core subjects within the standard K-12 curriculum, such as mathematics, science, language arts, or computational thinking, which has been steadily integrated into compulsory education worldwide, including Europe [3], the United States [30], and Brazil [4].
- *Self-Paced Adaptive Learning:* The system must include at least one of the following AI-driven components, ensuring real-time, student-centered adaptation: student modeling, knowledge tracing, knowledge graphs and representation, recommendation systems, personalized content generation, or automatic question generation.
- *Empirical Validation:* The study must include experimental results, user testing, or real-world deployment to ensure practical applicability and validation.
- *Publication Date:* The study must have been published from 2020 onwards to capture recent advancements in AI-based adaptive learning.
- *Open Access:* The study must be freely available to ensure reproducibility and accessibility for researchers and practitioners.

Exclusion Criteria Studies were excluded if they met any of the following conditions:

- *Non-K-12 Demographics:* Studies focused on higher education, vocational training, or adult learning were excluded, as this review is limited to K-12 education.
- *Supplemental Learning:* Studies on after-hours programs, tutoring, or teacher-assisted learning were excluded, as the focus is on AI-driven systems that function as primary instructors.
- *Non-Core Subjects:* Studies focusing on soft skills, emotional intelligence, physical education, foreign language learning, or other non-core curriculum topics were excluded to ensure relevance to standard academic subjects.

- *Non-Adaptive Performance Prediction:* Studies purely focused on predicting student outcomes on traditional assessments without providing real-time content adaptation were excluded.
- *Group-Based Adaptation:* Studies clustering students into groups for instruction rather than providing individualized, real-time adjustments were excluded, as they do not meet the criteria for hyper-personalized learning.
- *External Sensor Dependence:* Studies relying on IoT devices, wearables, or computer vision (e.g., eye tracking, gesture recognition) rather than analyzing student performance directly within the digital learning environment (e.g., computer, tablet, or mobile-based interactions) were excluded.
- *Demographic or Context-Specific Adaptation:* Studies primarily targeted at specific minority demographics (e.g., students with disabilities) or circumstantial adaptations (e.g., COVID-19 remote learning) were excluded to maintain generalizability.
- *Literature Reviews and Surveys:* Studies that solely provided systematic reviews, surveys, meta-analyses, or theoretical frameworks without implementing and testing an AI-driven adaptive learning system were excluded.

2.3 Screening Process

After executing the search, all retrieved articles were imported into a shared database. A single researcher conducted the title and abstract screening, followed by full-text review, using the inclusion and exclusion criteria described above.

At each stage, studies were evaluated for alignment with the review's focus on AI-driven, self-paced adaptive learning in K–12 education. Only studies meeting all criteria and providing empirical validation were considered for inclusion.

Although this review does not formally adhere to the PRISMA checklist, it follows a structured review protocol inspired by its emphasis on transparency and replicability. We applied a multi-stage screening process and clearly defined selection criteria to ensure methodological rigor. To visually summarize the study selection pipeline—from initial retrieval to final inclusion—we present a PRISMA-style flowchart in the following section.

3 Results

3.1 Study Selection

Our initial search yielded a total of 1031 articles—996 from the ACM Digital Library and 35 from IEEE Xplore—after filtering for publications from 2020 onward and applying the defined search string. During the title and abstract screening phase, 992 articles were excluded based on the predefined inclusion and exclusion criteria, as well as open access availability. As a result, 39 studies were retained for full-text review.

Following a comprehensive evaluation of the full texts, 8 studies were excluded for the following reasons:

- **Wang et al. (2024) [41]:** Excluded under the criterion of *Non-Adaptive Performance Prediction*. Although the system analyzes test results and visualizes performance data, it does not incorporate algorithm-driven personalization. All instructional decisions are made manually by teachers, and the platform functions solely as a dashboard. It lacks dynamic student modeling or AI-based content adaptation.
- **Hamim et al. (2022) [12]:** Excluded under the criterion of *Non-Adaptive Performance Prediction*. This study classifies students into static performance categories using boosting algorithms based on LMS data, but it does not implement any real-time, individualized adaptation of content or instruction.
- **Huang et al. (2023) [14]:** Excluded under the criterion of *Demographic or Context-Specific Adaptation*. The study focuses on mathematical question tagging using a Chinese-language domain ontology, limiting generalizability and applicability across broader K–12 populations.
- **Gao et al. (2020) [10]:** Excluded under the criterion of *Lack of Empirical Validation*. This article outlines conceptual steps for constructing course knowledge graphs but includes no implementation, testing, or real-world deployment.
- **Zhang and Tsai (2022) [46]:** Excluded under the criterion of *Lack of Methodological Transparency and Empirical Validation*. While proposing a multimedia system with AI, the paper lacks a defined modeling architecture, does not list input features, and omits any evaluation results.
- **Liu and Yuan (2024) [24]:** Excluded under the criterion of *Group-Based Adaptation*. The system provides group-based personalization strategies without real-time, individualized modeling or content sequencing.
- **Zhang and Huang (2024) [47]:** Excluded under the criterion of *Supplemental Learning*. The study evaluates motivational effects of adaptive quizzes but does not implement core components of AI-based instruction such as student modeling or adaptive sequencing.
- **Patel et al. (2022) [34]:** Excluded under the criterion of *Not Focused on AI-Based Instructional Adaptation*. The focus is on increasing access to tutoring systems via paper-digital integration, rather than developing adaptive learning mechanisms.

After this full-text screening phase, a total of **31 studies** were included in the final review. These studies form the analytical foundation of our synthesis on AI-driven, self-paced adaptive learning in K–12 education. The complete selection process is summarized in the PRISMA-style flowchart in Figure 1.

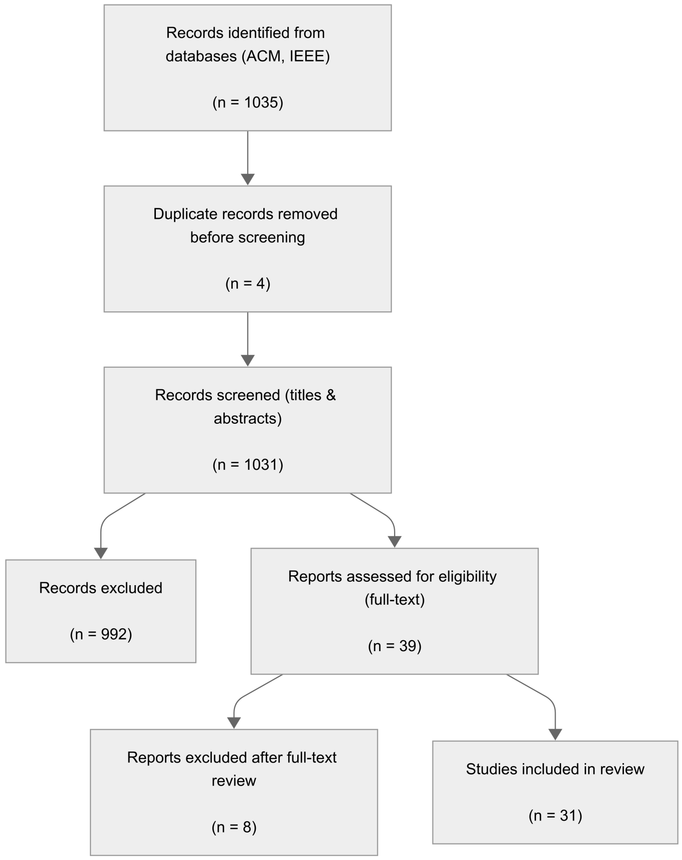

Fig. 1. PRISMA-style flowchart of the study selection process.

3.2 Overview of Selected Studies

To contextualize the scope and distribution of selected studies, Figure 2 provides an overview of their publication years and country origins. A total of 31 studies were included in this review, spanning from 2020 to 2025. As shown in the visualizations, the majority of contributions emerged in the years 2022 and 2024, reflecting a recent surge of interest in AI-driven, self-paced adaptive learning systems.

China stands out as the most prolific contributor, accounting for nearly half of all studies—often leading innovation in knowledge tracing, matrix factorization, and hybrid cognitive diagnosis. South Korea, the United States, and several European countries also contribute meaningfully to the field. While a few studies involved international collaborations, these were disaggregated to reflect each country's involvement individually.

This global distribution underscores the cross-regional interest in personalized K–12 education technologies, yet also highlights a concentration of research efforts in East and Southeast Asia. The temporal clustering around recent years confirms the rapidly evolving nature of the field and supports the timeliness of this review.

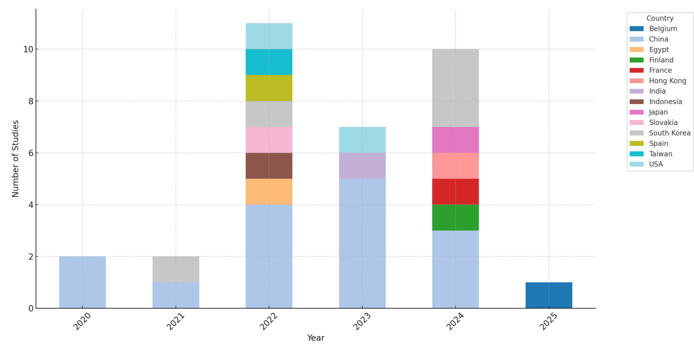

Fig. 2. Distribution of included studies by year and country

3.3 Student Modeling Approaches

Student modeling lies at the core of adaptive learning systems, enabling the estimation of learners' current knowledge, learning behaviors, and future performance. In this section, we classify student modeling approaches according to their primary methodological frameworks—such as knowledge tracing, cognitive diagnosis, and matrix factorization. However, these categories are not mutually exclusive. Many recent systems integrate techniques from multiple paradigms to combine their complementary strengths. For instance, [20] leverages monotonic attention from sequential modeling alongside item difficulty parameters inspired by Classical Test Theory (CTT), while [33] merges transformer-based architectures with interpretable constructs derived from Item Response Theory (IRT). These hybrid systems reflect a broader trend toward flexible, layered models that balance predictive power with pedagogical insight. As such, the taxonomy adopted here serves as an organizing lens rather than a rigid classification. .

Knowledge Tracing (KT) Knowledge tracing (KT) refers to the dynamic modeling of a student's evolving mastery over time, typically represented as a latent vector of concept proficiencies updated with each learning interaction. The goal is to predict future performance and support real-time personalization through accurate modeling of the learner's internal knowledge state. This section covers traditional, Bayesian, and deep learning-based KT models applied in K–12 adaptive learning systems.

Traditional and Bayesian KT Traditional KT models use explicit concept-to-question mappings (e.g., via Q-matrices) and parametric update rules grounded in learning and forgetting theories. For instance, the model proposed in [16] introduces Knowledge Proficiency Tracing (KPT), a transparent, equation-based framework that updates the concept mastery vector by integrating both Learning Curve Theory [29] and Forgetting Curve Theory [6]. The model considers each exercise as a vector over knowledge concepts and computes mastery changes after each attempt. An extended version, Exercise-Correlated KPT (EKPT) [16], incorporates concept-sharing across exercises. By modeling inter-exercise relationships, EKPT allows performance on one question to influence inferred mastery of related questions, thus improving prediction robustness in sparse datasets.

A more classical approach is represented by Bayesian Knowledge Tracing (BKT), which estimates the probability that a student knows a concept using four interpretable parameters: initial knowledge, learning rate, guess, and slip probabilities. A recent implementation of BKT [39] as the foundation for explainable recommendation systems in K–12 contexts shows the continuing relevance of probabilistic KT techniques in applications requiring transparency.

Transformer and Neural KT Recent years have seen a shift toward deep learning-based KT models, with particular emphasis on Transformer-style architectures that leverage attention mechanisms to model learning sequences. One prominent example is SAINT [28], which employs an encoder-decoder structure: the encoder captures the sequence of exercises, while the decoder models student responses using cross-attention. As sunch, SAINT introduces optimized attention layers and embedding strategies, resulting in improved prediction accuracy and robustness in modeling student learning trajectories.

CL4KT [21] builds on the Transformer backbone but introduces contrastive pretraining to better differentiate between similar and dissimilar learning histories. It uses separate Transformer encoders for question sequences and interaction sequences, and dynamically retrieves past responses to construct latent knowledge states, eliminating the need for external concept tags.

CRKT [33] adds interpretability by explicitly modeling why students chose particular options. It tracks not only selected answers but also compares them to unselected distractors to disentangle reasoning processes. CRKT further incorporates Item Response Theory (IRT) [7] principles to model the influence of concept difficulty and student ability.

Another approach, MonaCoBERT [20], adapts the BERT architecture for KT by introducing monotonic attention to model forgetting and convolutional attention to capture short-term behavior patterns. It enhances interpretability by integrating embeddings derived from Classical Test Theory (CTT) [5], representing question difficulty based on observed response data.

An additional Transformer-based system [42] combines self-attention mechanisms with temporal learning behavior (e.g., time on task, success rates) to infer a student's evolving knowledge state in human-robot interaction contexts.

Finally, DKVMN [38] represents a non-Transformer deep KT model. It leverages a memory-augmented architecture where each concept is a memory slot, updated through student responses via key-value attention. DKVMN enables flexible modeling of long-term dependencies and conceptual relationships, forming a bridge between neural KT and memory-based diagnostic systems.

These deep KT models offer powerful sequence modeling capabilities and greater flexibility, but often at the cost of reduced interpretability. Their rise reflects the growing importance of leveraging temporal structure and attention-based mechanisms to model student learning in real time.

Cognitive Diagnosis Models (CDM) Cognitive Diagnosis Models (CDMs) aim to estimate a student's mastery over fine-grained skills or concepts based on their response patterns. Unlike knowledge tracing, which models learning trajectories over time, CDMs focus on diagnosing the student's current knowledge state in a granular and interpretable way. These models are particularly useful in adaptive learning systems that target prerequisite reasoning, misconception identification, and multi-skill assessment. This section outlines recent CDM approaches categorized into probabilistic and graph-based methods and fuzzy or multi-skill aggregation models.

Probabilistic and Graph-Based CDM Bayesian and graph-based CDMs seek to quantify a learner's uncertainty and contextualize their performance across structured knowledge networks. For example, the BETA-CD framework [2] introduces a Bayesian meta-learned architecture that replaces the standard estimation of student ability (θ) with a hierarchical Bayesian layer capable of modeling prior knowledge and uncertainty. It is particularly well-suited to early-stage learners, as it adapts to sparse data and flags when confidence in a diagnosis is low—prompting further data collection.

A similarly probabilistic approach is found [40], which analyzes the co-occurrence of misconceptions in student responses. By using Bayesian probability and regression, the model identifies which incorrect beliefs tend to cluster together, enabling targeted instructional interventions.

Other recent models leverage structural knowledge representations. KI-EIR [23] implements a deep learning-based CDM known as NACD (Neural Attentive Cognitive Diagnosis), combining graph embeddings with student behavior modeling (e.g., guessing and slipping) to diagnose student understanding across concept networks. This allows for interpretable predictions in complex multi-concept scenarios.

An interesting hybrid approach is the PreferenceCD model by Jiang and Wang [17], which extends a traditional DINA framework by integrating student preferences into the diagnosis process. The model employs both a Q-matrix, manually linking exercises to concepts, and an M-matrix, which maps learning materials to concepts using a combination of teacher tagging and topic modeling (TF-IDF + LDA). While it remains binary and conjunctive in its reasoning, the system improves diagnostic accuracy by accounting for students' content engagement patterns — such as which texts they read most — thereby uncovering interest-aligned concept mastery. The PreferenceCD model significantly improves precision, RMSE, and MAE, outperforming traditional DINA and collaborative filtering baselines. It illustrates how behavioral data can enrich classical cognitive diagnosis, despite the added expert burden of constructing two aligned matrices.

Finally, the Graph Temporal Fusion Network (GFTN) [13] combines graph-based reasoning with temporal learning. Using a graph attention network (GAT) to model conceptual relationships and a gated recurrent unit (GRU) to capture learning progression over time, GFTN dynamically fuses both sources of information to track evolving student mastery. This makes it particularly useful for adaptive platforms that assess student knowledge across multiple sessions or timeframes.

Fuzzy and Multi-skill CDM Fuzzy logic and skill aggregation methods expand classical CDMs by modeling graded responses, multiple strategies, and compensatory skill interactions. The FuzzyCDF model, which incorporates the SI-GAM (Sugeno Integral-based Generalized Aggregation Method) [**?**]., supports both conjunctive and compensatory learning strategies while handling polytomous responses—such as partial correctness or degrees of agreement (e.g., "3 out of 5" or "strongly agree"). This allows for flexible combinations of mastered and unmastered skills, enabling more nuanced diagnosis in problem-solving tasks where students may follow different reasoning paths to reach correct or partially correct answers.

A simpler yet effective approach is [43], which uses local linear or weighted regression to model student mastery based on response trends. While not probabilistic or graph-based, it remains interpretable and adaptable, making it useful in systems requiring fast, lightweight cognitive diagnosis.

Together, these models demonstrate the growing diversity of CDM strategies in adaptive learning—spanning from probabilistic inference and deep graph reasoning to fuzzy logic and skill-weighted aggregation.

Matrix Factorization and Latent Embedding Models Matrix Factorization (MF) and latent embedding models learn vector representations of students and learning materials in a shared latent space, enabling personalized score prediction, content recommendation, and learning style inference. These models do not rely on predefined concept mappings or expert-crafted structures. Instead, they infer patterns of similarity and predictive relationships based on students' historical interactions with learning content, offering a scalable alternative to skill-tagged models such as KT and CDM.

A representative implementation is Wse-MF [37]. The system learns latent student vectors (P) and exercise vectors (Q) to predict performance scores using dot product operations. Wse-MF innovates on traditional MF by introducing a student-exercise-specific weighting strategy: it adjusts predictions based on inferred learning ability and question difficulty. This prevents overly optimistic predictions for low-ability students on difficult exercises by penalizing confidence where past performance data indicates risk.

In another application of MF, [?] integrates matrix factorization with constrained latent spaces to jointly estimate student learning styles and recommend tailored instructional resources. By embedding both students and content, the model aligns resource delivery with inferred preferences or needs—e.g., visual vs. text-based learners—allowing for more engaging and effective instruction. While it does not explicitly diagnose knowledge gaps, it contributes to the personalization layer by capturing individual learning tendencies.

A more complex approach is seen in RCES (Recommendation for Effective Standardized Exam Preparation) [27], a hybrid model that combines matrix factorization with sequential neural models for high-stakes test prediction. RCES models the interaction between students and test questions using MF, and then uses Bi-LSTM and Transformer architectures to read sequences of question-answer interactions over time. The model estimates a predicted TOEIC score for each learner, incorporating long-range behavioral patterns and attention-based reasoning. Additionally, an educational impact module tracks how feedback and performance updates influence expected outcomes, allowing RCES to not only predict but adapt over time.

Together, these models illustrate how matrix factorization and latent embeddings offer flexible, data-driven personalization mechanisms that can operate with minimal domain-specific annotations—an important advantage in large-scale K–12 deployments.

3.4 Domain Knowledge Representation

To personalize instruction effectively, adaptive learning systems must encode the structure of domain knowledge in a form that supports dynamic inference and recommendation. This structure may be embedded as an explicit mapping—such as a concept graph or Q-matrix—or as an implicit latent representation learned from student interaction data. While some systems rely on human-defined curricula, topic tags, or pedagogical taxonomies, others infer structure automatically from behavioral signals, content semantics, or sequential patterns.

In our review, we identified a recurring set of modeling strategies that allowed us to propose a two-dimensional classification framework. On one axis, we distinguish between *graph-based* systems—those that represent knowledge as interconnected concepts—and *non-graph-based* systems, which rely on concept tags, content labels, or vector representations. On the other axis, we differentiate between *expert-defined* structures and those *data-driven* or automatically inferred through machine learning techniques. This 2×2 taxonomy captures the fundamental design choices systems make when representing domain knowledge, highlighting trade-offs between interpretability, flexibility, and scalability.

The full taxonomy, including all 31 relevant studies, is presented in Table 1. Each cell of the matrix groups systems according to their structural assumptions and inference mechanisms, providing a foundation for the comparative analysis that follows.

Table 1. Taxonomy of domain knowledge representation in selected adaptive learning systems. Full references are listed in the bibliography.

Representation Type	Expert-Defined	Data-Driven
Graph-Based	– [45] – [33] (expert-defined concept maps) – [25] – [23]	– [1] – [44] – [13] – [11] – [33] (statistical inference) – [28]
Non-Graph-Based	– [37] – [16] – [39] – [18] – [43] – [42] – [48] – [31] – [26] – [32] – [22] – [49] – [35] – [17]	– [20] – [15] – [21] – [28] – [37]

Graph-Based and Expert-Defined Systems in this category leverage *curriculum-aligned concept graphs* or teacher-authored maps to explicitly encode dependencies between concepts. These structures often reflect pedagogical hierarchies or prerequisite relationships.

The study by [45] constructs a concept graph from curriculum documents using BiLSTM+CRF, with teachers manually defining node centrality weights to prioritize concept importance. Similarly, [23] introduces a multi-level knowledge graph (subject, concept, task) and computes concept importance using five structural and semantic features. [33] partially relies on expert-defined concept maps to contextualize student misconceptions and uses contrastive learning to capture nuanced reasoning differences. Meanwhile, [25] applies a predefined concept structure based on the official Taiwanese curriculum, anchoring adaptivity in established educational standards. These systems offer strong interpretability but are limited by scalability and potential biases in expert-defined hierarchies.

Graph-Based and Data-Driven This quadrant includes systems that *infer concept graphs automatically* from raw data, such as behavioral logs, content metadata, or semantic embeddings. These approaches capture emergent relationships and are more scalable across domains.

For example, [1] constructs a multi-level semantic graph from open educational resources using Wikification and DBpedia enrichment, while [44] applies RoBERTa to extract semantic features from course descriptions and behavioral data. [13] uses graph attention networks (GATs) to infer relationships between knowledge points and question items. [11] builds a performance-concept map (PCM) from student response patterns, creating a bipartite graph that connects students and knowledge points. Even [33], when using co-occurrence statistics rather than expert maps, fits this quadrant. Notably, [28] also learns soft dependencies between exercises using attention weights—functioning as a form of implicit graph inference. These systems support flexible modeling but often trade off interpretability and domain alignment.

Non-Graph-Based and Expert-Defined Systems in this category use *structured but non-relational representations* of knowledge, such as expert-authored Q-matrices, topic labels, or taxonomies. They offer reliable curriculum alignment without modeling inter-concept relationships.

Many rely on Q-matrices, like [37], which combines matrix factorization with expert-defined mappings between exercises and skills, and [16], which extends this by modeling concept similarity through co-tagging. [39] integrates Q-matrix tags into a probabilistic knowledge tracing system for explainable recommendations. Other systems use rule-based or hierarchical topic labels—e.g., [18], [43], [35], and [31]—without encoding concept dependencies. [49] classifies math problems into categories for generation and extraction but does not construct a concept graph.

A particularly notable contribution in this category is the PreferenceCD model by Jiang and Wang [17], which integrates a traditional DINA-based cognitive diagnosis framework with students' learning material preferences. It employs both a Q-matrix—manually tagged by teachers to map exercises to concepts—and an M-matrix linking materials to concepts. While the M-matrix is partially expert-labeled, it also incorporates topic modeling techniques such as TF-IDF and LDA to enhance semantic alignment. This dual-matrix structure allows the system to reason about both formal assessments and informal reading behaviors, enriching student diagnosis with insights drawn from preferred materials. Despite requiring a high expert burden for tagging, it exemplifies a hybrid approach that maintains curriculum alignment while introducing semi-automated enhancements to domain modeling.

Non-Graph-Based and Data-Driven This quadrant includes systems that *learn internal, latent representations* of knowledge without relying on expert structures. These models, typically powered by transformers, matrix factorization, or contrastive learning, excel in flexibility and predictive performance.

[20] uses monotonic attention to model forgetting and concept interaction, producing latent dependency graphs as a byproduct of learned embeddings. [15] learns concept mastery vectors directly from response patterns, while [21] uses contrastive pretraining to differentiate student trajectories without predefined concept tags. [28] employs QKV attention mechanisms to align student responses with latent concept structures. [37] appears here as well, given its dual role in learning dense latent vectors in addition to its Q-matrix foundation. These systems push the boundaries of adaptivity, though they often lack transparency for educators and rely on large datasets for training stability.

4 Personalization Mechanisms

While student modeling provides a foundation for understanding what learners know, personalization mechanisms determine how that knowledge is used to adapt instruction. These mechanisms define how educational systems respond to real-time inputs and long-term learner profiles to tailor the sequence, format, difficulty, or generation of content.

Based on our review, we identified four primary types of personalization in AI-based adaptive learning systems:

- **Learning Path Recommendation:** Selecting or optimizing the sequence of topics or instructional units a student should follow.
- **Modality and Resource Personalization:** Adapting the type or format of learning content (e.g., visual vs. textual, video vs. interactive).
- **Exercise Sequencing and Difficulty Adjustment:** Recommending problem sets and dynamically adjusting difficulty based on performance or learner choice.
- **Personalized Content Generation:** Creating entirely new exercises or materials using generative AI or procedural logic.

The following subsections present each category in detail, including representative systems, technical approaches, and intended pedagogical effects.

Learning Path Recommendation Several systems in our review focused on optimizing the sequence of instructional units or topics a student encounters. Some approaches remain curriculum-centric: for example, [45] uses a multi-iteration optimization algorithm (MI-SA-PSO) to recommend paths across a concept graph based on node centrality and path diversity. However, this system provides the same path to all students unless manually modified. Similarly, [25] recommends pre-validated learning paths aligned with the official national curriculum, without dynamic personalization.

More adaptive approaches integrate real-time learner data. [42] combines a Transformer encoder with a reinforcement learning (RL) agent that recommends materials based on evolving student behavior, allowing for personalized path optimization. [39] applies a Bayesian Knowledge Tracing model to identify

"Goldilocks" quizzes—those with predicted correctness near 50%—balancing difficulty and curriculum alignment. It also prioritizes foundational content through ordering weights and uses model parameters (e.g., guess/slip) to generate personalized motivational explanations. Finally, [22] personalizes video recommendations using longitudinal IRT-based ability tracking, guiding students through review or advancement in line with their zone of proximal development.

Modality and Resource Personalization This category includes systems that adapt the format or type of instructional content to student preferences or engagement patterns. [35] applies a Deep Q-Network (DQN) to select presentation modalities—visual, auditory, read/write, or kinesthetic (VARK)—based on interaction behavior. The RL agent receives a reward signal based on learning gains, time efficiency, and reduced hint usage, and was pre-trained using a simulated student model over one million episodes.

[32] predicts student learning styles using matrix factorization and recommends resources accordingly, offering a collaborative filtering-based approach to modality adaptation. Meanwhile, [18] combines Random Forest classification with DeepAFM to recommend materials (e.g., videos, exercises) based on behavior metrics like time-on-task and submission frequency. The system adapts gamified rewards to align with inferred effort, making the personalization not only about content, but also motivational reinforcement.

Finally, [17] personalizes instructional material by modeling students' preferred learning resources based on behavior logs and reading frequency. Rather than relying on predefined learning styles, they infer individual preferences using both direct methods (topic modeling of read materials) and indirect clustering (similarity between students). This allows the system to recommend the most relevant content—both for performance prediction and engagement.

Exercise Sequencing and Difficulty Adjustment Exercise-level personalization plays a central role in several adaptive platforms. [35] uses rule-based logic to scaffold task difficulty based on penalized scores—adjusted for hint usage and failed attempts. Students begin at a level determined by pre-tests, and the system promotes or demotes difficulty dynamically.

[16] recommends semantically similar exercises based on shared concept tags and exercise proximity, improving robustness in sparse data contexts. [23] selects questions that are both informative (diagnostically useful) and representative (conceptually broad), enabling efficient progression. [31] adds a layer of transparency and learner agency by letting students adjust exercise difficulty through a slider, offering predicted mastery gains via visual "what-if" charts and contextual motivational feedback. Lastly, [26] allows students to specify the concept and difficulty level of programming exercises—offering lightweight, user-driven personalization without adaptive modeling.

Personalized Content Generation While most systems focus on selecting pre-existing material, a few generate new content on demand. [42] employs GANs to create learning materials (e.g., questions, visual aids) aligned with the learner's current level and style. A discriminator ensures quality and alignment with educational objectives, allowing content to evolve in response to performance and engagement.

[49] leverages GPT-3 to generate math word problems across topics, using classification and equation extraction to create problems customized by category and complexity. These generative systems offer high flexibility but also raise challenges around quality assurance, pedagogical alignment, and interpretability.

4.1 Implementation and Evaluation

To assess the effectiveness and applicability of adaptive learning systems in K–12 contexts, we analyzed the implementation and evaluation methodologies of the 31 selected studies. This section synthesizes key trends in the subject domains addressed, types of evaluation performed, and reported outcomes in learning performance, engagement, or system capabilities.

Subject Domains While not every study explicitly stated its instructional subject, a substantial majority focused on mathematics and English language learning. Mathematics appeared most frequently, with systems such as CRKT [33], KI-EIR [23], Wse-MF [37], GFTN [13], and EKPT [16] targeting topics ranging from basic algebra to advanced diagnostics. English instruction was also common, especially in models evaluated using datasets like the Michigan English Proficiency Exam or EdNet [20, 2]. A smaller number of systems addressed computer science and programming education, including generative exercise platforms [26] and behavioral modeling tools [18]. Isolated studies focused on physics [40] or cross-disciplinary open educational resources [1].

Evaluation Methodologies

Dataset-Based Evaluation Most studies validated their models using real-world educational datasets, such as ASSISTments, EdNet, Algebra05/06, and TOEIC preparation logs. These evaluations typically focused on predictive accuracy using metrics like AUC, RMSE, or ranking precision [21, 20, 2, 28, 37]. Several systems, including CL4KT [21], MonaCoBERT [20], and BETA-CD [2], conducted extensive ablation studies to isolate the contributions of specific model components.

Case Studies and In-Situ Deployments Twelve studies included exploratory or qualitative deployments in real classroom settings. For instance, the DMP_AI system was piloted in eight schools with staff surveys but no student testing [44], while the Taiwanese platform [25] and the AI-based VARK intervention [35] included pre- and post-tests but lacked control groups. User-centered studies like [31] explored student preferences via think-aloud sessions and feedback on motivational interfaces.

Controlled Experiments Only two studies reported true experimental designs. The RCES study [27] conducted an A/B test with 1,713 TOEIC students across four recommendation systems, showing clear performance gains with adaptive personalization. Similarly, the Algebra Nation study [22] deployed a randomized controlled trial with 2,995 students across 42 Florida schools, demonstrating significant learning gains from personalized video sequencing.

Reported Outcomes Despite the diversity of systems and evaluation settings, several trends emerged across studies:

- **Improved Predictive Accuracy:** Deep learning-based KT systems consistently outperformed traditional baselines. For example, CL4KT [21] and SAINT [28] achieved the highest AUCs across multiple datasets, particularly excelling in sparse data environments.
- **Enhanced Learning Gains:** Systems with real-world deployments often reported positive, though sometimes modest, improvements. The Algebra Nation RCT found a 0.33 standard deviation gain in post-test scores [22], while the VARK-based intervention increased learning efficiency by over 150% [35].
- **User Engagement and Satisfaction:** Studies like [31] and [18] highlighted the motivational benefits of gamified elements and student-controlled interfaces, even when direct learning gains were not measured.
- **Robustness to Data Scarcity:** Systems such as EKPT [16] and BETA-CD [2] demonstrated stability and accurate diagnostic performance even with limited training data, aided by meta-learning or exercise correlation modeling.
- **Interpretability and Transparency:** CRKT [33] and BKT-based systems like [39] offered human-readable explanations, helping educators and students understand system behavior—a notable advantage in K–12 contexts where trust and pedagogy alignment matter.
- **Generative Potential:** GPT-3-based systems [49] and GAN-based material generation [42] showed promising capabilities in customizing problems, though concerns about alignment and validation remain.

Together, these results suggest that while many adaptive systems are still in experimental or early-stage deployment phases, their performance—especially in predictive modeling and personalization—shows clear promise. However, few systems are robustly validated through randomized controlled trials or longitudinal studies, limiting our understanding of long-term impact and scalability.

5 Discussion

5.1 Trade-offs Between Automation and Manual Labeling

Despite the rise of powerful data-driven modeling techniques, many adaptive learning systems still rely heavily on manual labeling strategies such as Q-matrices, topic taxonomies, and curriculum-aligned structures [37, 16, 39, 18, 33, 25, 49, 23, 35, 26]. These approaches improve interpretability and alignment with expert knowledge but introduce scalability challenges. For instance, systems may misinterpret surface-level student behavior without deeper semantic modeling—such as mistaking difficulty with word problems for a lack of algebra skills when the issue may be reading comprehension.

In contrast, models like [21] and [44] infer structure automatically. [21] eliminates the need for concept tags entirely, using sequence augmentation and contrastive learning. [44] applies semantic feature extraction with RoBERTa and builds heterogeneous student-resource graphs to infer relationships. PreferenceCD [17], in turn, highlights the tension between manual and automated domain modeling. While the system relies on expert-defined Q- and M-matrices, the integration of LDA and TF-IDF in tagging learning materials offers a semi-automated compromise—balancing scalability and semantic alignment. These methods scale more easily and adapt across domains but sacrifice transparency, making them harder to debug or justify in educational settings.

5.2 The Black-Box Problem vs. Explainability in Adaptive Systems

The increasing complexity of adaptive systems often results in black-box models that perform well but lack interpretability. Examples include [21, 1, 42, 2], which rely on deep networks without explicit, human-readable explanations. This can be problematic in K–12 environments where trust, accountability, and the ability to provide formative feedback are crucial.

By contrast, explainable systems like [16, 37, 39, 9] offer intrinsic transparency. [20] visualizes attention scores over past interactions; [13] highlights graph node influences over time. [33] stands out by analyzing distractors to model student misconceptions, while [31] investigates the pedagogical impact of interface-level explanation types like "why" and "what-if". Notably, PreferenceCD [17] offers a transparent way to interpret performance anomalies through interest-driven behavior. For instance, students may excel on a topic they've informally studied through preferred materials, even if their formal mastery remains limited—an insight that traditional CDMs may miss. These efforts align AI recommendations with teacher expectations and support student agency.

5.3 The Rise of Deep Learning and Hybrid Architectures

Transformer-based and hybrid deep learning models now dominate the field of adaptive education. Models such as [28, 21, 20, 27, 23, 2] achieve strong performance across sparse and noisy datasets by integrating multiple architectural innovations, including memory networks, attention layers, and graph-temporal reasoning modules. For example, [13] fuses GAT and GRU to capture both structural and temporal dimensions of learning.

However, this trend increases opacity. Some systems attempt hybrid solutions to bridge the gap between accuracy and interpretability—such as combining explainable decision trees with deep content recommenders, as seen in [18], or integrating reinforcement learning into rule-based personalization frameworks, like in [35].

5.4 Beyond Mastery: Modeling Misconceptions, Preconceptions, and Importance

Accurate modeling of student knowledge goes beyond binary mastery. Several systems address this by explicitly modeling misconceptions. [40] uses Bayesian inference to identify and categorize student preconceptions. [33] analyzes unchosen answers to reveal reasoning errors, enhancing the granularity of diagnosis.

Some systems focus on learning priorities rather than deficiencies. [11] estimates the importance of concepts for future learning, while [18] uses sentiment and behavioral NLP to detect confusion and misunderstanding from student discussion posts. These innovations move adaptive learning closer to the formative, diagnostic role played by effective teachers.

5.5 Modeling Learning Curves, Forgetting, and Cognitive Variability

Student knowledge changes over time—not only through learning, but also through forgetting. [16] explicitly encodes this with dual learning and forgetting curves, improving both predictive performance and interpretability. [20] applies monotonic attention to simulate memory decay, modeling how recent interactions weigh more heavily in student state estimation.

Meanwhile, [21] introduces cognitive variability through data augmentation, simulating different learning paths and attention patterns. These methods allow systems to be more resilient to incomplete or noisy learning histories—an important characteristic in real-world educational deployments.

5.6 Motivational Mechanisms and Engagement Estimation

Only a few systems in the review directly address learner motivation and engagement. [22] tracks video follow-through rates as a proxy for motivation, revealing subgroup differences in responsiveness. [31] embeds motivational framing into feedback messages and allows students to control difficulty, improving both agency and engagement.

Systems like [35] adapt content modality in real time using reinforcement learning—responding dynamically to learner behavior rather than static preferences. These approaches represent early steps toward systems that model affective states and personalize not only what is taught, but how and when it is delivered.

5.7 The Fragility of Learning Styles as a Personalization Mechanism

Despite longstanding criticism of learning styles in education research, several systems still use them as a basis for personalization [35, 32, 43]. While most employ learning styles as an initial signal rather than a rigid framework, their empirical grounding remains weak.

For instance, [35] uses a VARK questionnaire to initialize content presentation but rapidly updates based on real-time performance. This dynamic adjustment is more defensible, yet interpretation of learning gains in such studies must be cautious—especially when no control groups or ablation comparisons are included.

5.8 Technical Transparency and Reporting Gaps

Several papers reviewed lacked the technical detail required for reproducibility or critical evaluation. [44, 48, 31] omitted key implementation details, including model architecture, training procedures, or hyperparameter settings. In some cases, only qualitative outcomes were reported, with no control group or statistical testing [43, 26].

This lack of transparency limits the ability to assess causal impact, reproduce findings, or scale systems to new contexts. As adaptive learning research matures, standardized reporting—covering data sources, evaluation design, and model interpretability—should become a baseline requirement for publication and policy relevance.

6 Conclusion

This review synthesized 31 recent studies on AI-driven, self-paced adaptive learning in K–12 education. While progress has been made in student modeling and content personalization, several gaps remain.

Many systems still rely on manually defined Q-matrices and topic labels, which limit scalability across subjects and languages. Future research should prioritize automated methods for extracting knowledge structures directly from curricula, combining semantic modeling with curriculum-aligned logic. Deep learning models continue to outperform classical approaches, especially in sparse data contexts, but often sacrifice interpretability. Building explainable models that align with classroom needs remains an urgent challenge.

Although some systems use learning styles as personalization anchors, most do not empirically validate their effects. When used, learning styles should act as soft priors, updated dynamically based on student behavior rather than fixed questionnaires.

A key limitation across the field is the lack of robust real-world experimentation. Only two studies in this review used randomized controlled trials. Broader deployment and rigorous testing are essential to assess long-term impact on learning.

Finally, the rise of large language models opens new opportunities for content generation, but raises questions about quality, alignment, and accountability. These tools should be integrated cautiously, with teacher oversight and clear evaluation standards.

In the end, the goal is not just to model what students know—but to design systems that learn how they learn, and explain themselves along the way

References

1. Bazouzi, A., Le Capitaine, H., Miklos, Z., Foursov, M.: Precedability prediction between open educational resources. In: Proceedings of the 2024 International Conference on Information Technology for Social Good. p. 386–393. GoodIT '24, ACM (Sep 2024). https://doi.org/10.1145/3677525.3678686, http://dx.doi.org/10.1145/3677525.3678686
2. Bi, H., Chen, E., He, W., Wu, H., Zhao, W., Wang, S., Wu, J.: Beta-cd: A bayesian meta-learned cognitive diagnosis framework for personalized learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 5018–5026 (2023). https://doi.org/10.1609/aaai.v37i4.25629, https://ojs.aaai.org/index.php/AAAI/article/view/25629
3. Bocconi, S., Inamorato dos Santos, A., Chioccariello, A., Cachia, R., Kampylis, P., Giannoutsou, N., Dagienė, V., Punie, Y., Wastiau, P., Engelhardt, K., Earp, J., Horvath, M., Jasutė, E., Malagoli, C., Masiulionytė-Dagienė, V., Stupurienė, G.: Reviewing computational thinking in compulsory education: State of play and practices from computing education. Research report, European Commission: Joint Research Centre, Luxembourg (2022). https://doi.org/10.2760/126955, https://data.europa.eu/doi/10.2760/126955
4. Brasil, Ministério da Educação: Base nacional comum curricular (2018), http://basenacionalcomum.mec.gov.br
5. Brookhart, S.M., Nitko, A.J.: Educational Assessment of Students. Pearson, Boston, MA, 8th edn. (2018)
6. Ebbinghaus, H.: Memory: A contribution to experimental psychology. Teachers College Press (1913). https://doi.org/10.1037/10011-000, http://dx.doi.org/10.1037/10011-000
7. Embretson, S.E., Reise, S.P.: Item Response Theory. Psychology Press (Sep 2013). https://doi.org/10.4324/9781410605269, http://dx.doi.org/10.4324/9781410605269
8. Ferreira, S.G., Ribeiro, G., Tafner, P.: Abandono e evasão escolar no brasil. Nota Técnica NT 2022/1, Instituto Mobilidade e Desenvolvimento Social (Imds), Rio de Janeiro (june 2022), https://imdsbrasil.org/publicacao/abandono-e-evasao-escolar-no-brasil/
9. Fumanal-Idocin, J., Takáč, Z., Horanská, , da Cruz Asmus, T., Dimuro, G., Vidaurre, C., Fernandez, J., Bustince, H.: A generalization of the sugeno integral to aggregate interval-valued data: An application to brain computer interface and social network analysis. Fuzzy Sets and Systems **451**, 320–341 (Dec 2022). https://doi.org/10.1016/j.fss.2022.10.003, http://dx.doi.org/10.1016/j.fss.2022.10.003
10. Gao, J., Wang, L., Xu, F.: Research on the construction of course knowledge graph of high school information technology. In: 2020 International Conference on Artificial Intelligence and Education (ICAIE). pp. 211–215 (2020). https://doi.org/10.1109/ICAIE50891.2020.00056
11. Guo, H., Yu, X., Wang, X., Guo, L., Xu, L., Lu, R.: Discovering knowledge-point importance from the learning-evaluation data. International Journal of Distance Education Technologies (IJDET) **20**(1), 1–20 (2022)

12. Hamim, T., Benabbou, F., Sael, N.: Student profile modeling using boosting algorithms. Int. J. Web-Based Learn. Teach. Technol. **17**(5), 1–13 (Sep 2022). https://doi.org/10.4018/IJWLTT.20220901.oa4, https://doi.org/10.4018/IJWLTT.20220901.oa4
13. Huang, M., Wei, T.: Gtfn: Knowledge tracing model based on graph temporal fusion networks. Int. J. Data Warehous. Min. **20**(1), 1–17 (May 2024). https://doi.org/10.4018/IJDWM.345406, https://doi.org/10.4018/IJDWM.345406
14. Huang, T., Hu, S., Lin, K., Yang, H., Zhang, H., Song, H., Lv, Z.: Sequence generation model integrating domain ontology for mathematical question tagging. ACM Transactions on Asian and Low-Resource Language Information Processing (2023)
15. Huang, T., Yang, H., Li, Z., Xie, H., Geng, J., Zhang, H.: A dynamic knowledge diagnosis approach integrating cognitive features. IEEE Access **9**, 116814–116829 (2021). https://doi.org/10.1109/ACCESS.2021.3105830
16. Huang, Z., Liu, Q., Chen, Y., Wu, L., Xiao, K., Chen, E., Ma, H., Hu, G.: Learning or forgetting? a dynamic approach for tracking the knowledge proficiency of students. ACM Trans. Inf. Syst. **38**(2) (Feb 2020). https://doi.org/10.1145/3379507, https://doi.org/10.1145/3379507
17. Jiang, P., Wang, X.: Preference cognitive diagnosis for student performance prediction. IEEE Access **8**, 219775–219787 (2020). https://doi.org/10.1109/ACCESS.2020.3042775
18. Jiao, X., Yu, X., Peng, H., Zhang, X.: A smart learning assistant to promote learning outcomes in a programming course. Int. J. Softw. Sci. Comput. Intell. **14**(1), 1–23 (Nov 2022). https://doi.org/10.4018/IJSSCI.312557, https://doi.org/10.4018/IJSSCI.312557
19. Kabudi, T., Pappas, I., Olsen, D.H.: Ai-enabled adaptive learning systems: A systematic mapping of the literature. Computers and Education: Artificial Intelligence **2**, 100017 (2021). https://doi.org/10.1016/j.caeai.2021.100017, https://www.sciencedirect.com/science/article/pii/S2666920X21000114
20. Lee, U., Park, Y., Kim, Y., Choi, S., Kim, H.: Monacobert: Monotonic attention based convbert fornbsp;knowledge tracing. In: Generative Intelligence and Intelligent Tutoring Systems: 20th International Conference, ITS 2024, Thessaloniki, Greece, June 10–13, 2024, Proceedings, Part II. p. 107–123. Springer-Verlag, Berlin, Heidelberg (2024). https://doi.org/10.1007/978-3-031-63031-6_10, https://doi.org/10.1007/978-3-031-63031-6_10
21. Lee, W., Chun, J., Lee, Y., Park, K., Park, S.: Contrastive learning for knowledge tracing. In: Proceedings of the ACM Web Conference 2022. p. 2330–2338. WWW '22, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3485447.3512105, https://doi.org/10.1145/3485447.3512105
22. Leite, W.L., Kuang, H., Shen, Z., Chakraborty, N., Michailidis, G., D'Mello, S., Xing, W.: Heterogeneity of treatment effects of a video recommendation system for algebra. In: Proceedings of the Ninth ACM Conference on Learning @ Scale. p. 12–23. L@S '22, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3491140.3528275, https://doi.org/10.1145/3491140.3528275
23. Li, L., Wang, Z.: Knowledge graph-enhanced intelligent tutoring system based on exercise representativeness and informativeness. International Journal of Intelligent Systems **2023**(1) (Jan 2023). https://doi.org/10.1155/2023/2578286, http://dx.doi.org/10.1155/2023/2578286
24. Liu, B., Yuan, D.: Research on personalized teaching strategies based on learner profiles in a blended learning environment. Int. J. Inf. Commun.

Technol. Educ. **20**(1), 1–25 (Jul 2024). https://doi.org/10.4018/IJICTE.346823, https://doi.org/10.4018/IJICTE.346823

25. Liu, T.C.: A case study of the adaptive learning platform in a taiwanese elementary school: Precision education from teachers' perspectives. Education and Information Technologies **27**(5), 6295–6316 (Jun 2022). https://doi.org/10.1007/s10639-021-10851-2, https://doi.org/10.1007/s10639-021-10851-2
26. Logacheva, E., Hellas, A., Prather, J., Sarsa, S., Leinonen, J.: Evaluating contextually personalized programming exercises created with generative ai. In: Proceedings of the 2024 ACM Conference on International Computing Education Research - Volume 1. p. 95–113. ICER '24, Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3632620.3671103, https://doi.org/10.1145/3632620.3671103
27. Loh, H., Shin, D., Lee, S., Baek, J., Hwang, C., Lee, Y., Cha, Y., Kwon, S., Park, J., Choi, Y.: Recommendation for effective standardized exam preparation. In: LAK21: 11th International Learning Analytics and Knowledge Conference. p. 397–404. LAK21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3448139.3448177, https://doi.org/10.1145/3448139.3448177
28. Lu, G., Niu, K., Peng, X., Zhou, Y., Zhang, K., Tai, W.: Self-kt: Self-attentive knowledge tracing with feature fusion pre-training in online education. In: 2024 International Joint Conference on Neural Networks (IJCNN). pp. 1–8 (2024). https://doi.org/10.1109/IJCNN60899.2024.10651418
29. Newell, A., Rosenbloom, P.S.: Mechanisms of skill acquisition and the law of practice. In: Cognitive skills and their acquisition, pp. 1–55. Psychology Press (2013)
30. Next Generation Science Standards (NGSS): Next generation science standards search, https://www.nextgenscience.org/search-standards
31. Ooge, J., Vanneste, A., Szymanski, M., Verbert, K.: Designing visual explanations and learner controls to engage adolescents in ai-supported exercise selection. p. 1–12. LAK '25, Association for Computing Machinery, New York, NY, USA (2025). https://doi.org/10.1145/3706468.3706470, https://doi.org/10.1145/3706468.3706470
32. Pardamean, B., Suparyanto, T., Cenggoro, T.W., Sudigyo, D., Anugrahana, A.: Ai-based learning style prediction in online learning for primary education. IEEE Access **10**, 35725–35735 (2022). https://doi.org/10.1109/ACCESS.2022.3160177
33. Park, S., Lee, D., Park, H.: Enhancing knowledge tracing with concept map and response disentanglement. Know.-Based Syst. **302**(C) (Oct 2024). https://doi.org/10.1016/j.knosys.2024.112346, https://doi.org/10.1016/j.knosys.2024.112346
34. Patel, N., Thakkar, M., Rabadiya, B., Patel, D., Malvi, S., Sharma, A., Lomas, D.: Equitable access tonbsp;intelligent tutoring systems through paper-digital integration. In: Intelligent Tutoring Systems: 18th International Conference, ITS 2022, Bucharest, Romania, June 29 – July 1, 2022, Proceedings. p. 255–263. Springer-Verlag, Berlin, Heidelberg (2022). https://doi.org/10.1007/978-3-031-09680-8_24, https://doi.org/10.1007/978-3-031-09680-8$_2$4
35. Sayed, W.S., Noeman, A.M., Abdellatif, A., Abdelrazek, M., Badawy, M.G., Hamed, A., El-Tantawy, S.: Ai-based adaptive personalized content presentation and exercises navigation for an effective and engaging e-learning platform. Multimedia Tools and Applications **82**(3), 3303–3333 (Jun 2022). https://doi.org/10.1007/s11042-022-13076-8, http://dx.doi.org/10.1007/s11042-022-13076-8
36. Skinner, B.F.: Teaching machines. Science **128**(3330), 969–977 (1958). https://doi.org/10.1126/science.128.3330.969, https://www.science.org/doi/10.1126/science.128.3330.969

37. Sun, X., Li, B., Sutcliffe, R., Gao, Z., Kang, W., Feng, J.: Wse-mf: A weighting-based student exercise matrix factorization model. Pattern Recognition **138**, 109285 (Jun 2023). https://doi.org/10.1016/j.patcog.2022.109285, http://dx.doi.org/10.1016/j.patcog.2022.109285
38. Sun, X., Zhao, X., Li, B., Ma, Y., Sutcliffe, R., Feng, J.: Dynamic key-value memory networks with rich features for knowledge tracing. IEEE Transactions on Cybernetics **52**(8), 8239–8245 (2022). https://doi.org/10.1109/TCYB.2021.3051028
39. Takami, K., Flanagan, B., Dai, Y., Ogata, H.: Evaluating the effectiveness of bayesian knowledge tracing model-based explainable recommender. Int. J. Distance Educ. Technol. **22**(1), 1–23 (Feb 2024). https://doi.org/10.4018/IJDET.337600, https://doi.org/10.4018/IJDET.337600
40. Wang, J., Liang, K.: A cognitive diagnosis method in adaptive learning system based on preconceptions. Scientific Programming **2022**, 1–10 (Jan 2022). https://doi.org/10.1155/2022/5011804, http://dx.doi.org/10.1155/2022/5011804
41. Wang, L., Wang, X., Fan, M.: The construction and application of precise teaching mode based on educational big data. In: 2024 International Symposium on Educational Technology (ISET). pp. 411–415 (2024). https://doi.org/10.1109/ISET61814.2024.00087
42. Xiong, L., Chen, Y., Peng, Y., Ghadi, Y.Y.: Improving robot-assisted virtual teaching using transformers, gans, and computer vision. J. Organ. End User Comput. **36**(1), 1–32 (Jan 2024). https://doi.org/10.4018/JOEUC.336481, https://doi.org/10.4018/JOEUC.336481
43. Yang, H., Shankar, A., S., V.: Artificial intelligence-enabled interactive system modeling for teaching and learning based on cognitive web services. International Journal of e-Collaboration **19**(2), 1–18 (Jan 2023). https://doi.org/10.4018/ijec.316655, http://dx.doi.org/10.4018/ijec.316655
44. Yang, Z.Q., Cao, J., Li, X., Wang, K., Zheng, X., Poon, K.C.F., Lai, D.: DMP$_AI : AnAI-AidedK-12SystemforTeachingandLearninginDiverseSchools, p.117\check{}130.SpringerNatureSingapore(2024).https://doi.org/10.1007/978-981-97-4442-8_9, http://dx.doi.org/10.1007/978-981-97-4442-8_9$
45. Zhang, J., Xia, R., Wang, Q.: Design of data-driven learning path based on knowledge graph and tracing model. In: 2023 IEEE International Conference on High Performance Computing Communications, Data Science Systems, Smart City Dependability in Sensor, Cloud Big Data Systems Application (HPCC/DSS/SmartCity/DependSys). pp. 813–820 (2023). https://doi.org/10.1109/HPCC-DSS-SmartCity-DependSys60770.2023.00118
46. Zhang, Y., Tsai, S.B.: Development and application of artificial intelligence multimedia technology based on big data. Mob. Inf. Syst. **2022** (Jan 2022). https://doi.org/10.1155/2022/2073091, https://doi.org/10.1155/2022/2073091
47. Zhang, Z., Huang, X.: Exploring the impact of the adaptive gamified assessment on learners in blended learning. Education and Information Technologies **29**(16), 21869–21889 (May 2024). https://doi.org/10.1007/s10639-024-12708-w, http://dx.doi.org/10.1007/s10639-024-12708-w
48. Zhu, J., Liao, Y., Lu, D.: Design and optimization of personalized education system based on intelligent algorithms. Procedia Computer Science **243**, 514–522 (2024). https://doi.org/10.1016/j.procs.2024.09.063, http://dx.doi.org/10.1016/j.procs.2024.09.063
49. Zong, M., Krishnamachari, B.: Solving math word problems concerning systems of equations with gpt-3. In: Proceedings of the Thirty-Seventh AAAI Confer-

ence on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence. AAAI'23/IAAI'23/EAAI'23, AAAI Press (2023). https://doi.org/10.1609/aaai.v37i13.26896, https://doi.org/10.1609/aaai.v37i13.26896