2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte

Paper ID #51021

Personalization Strategies in Large Language Model–Based Educational Systems: A Scoping Review

Onil Morshed, Gettysburg College

Onil Morshed is a senior at Gettysburg College double majoring in Computer Science and Mathematics.

VINCENT OLUWASETO FAKIYESI, University of Georgia

Vincent Oluwaseto Fakiyesi is a doctoral student in the Engineering Education program at the University of Georgia. He holds a bachelor's degree in Chemical Engineering and is interested in advancing assessment practices to better understand student learning. His research focuses on cognitive diagnostic modeling, student misconceptions, and adaptive assessment systems. He applies diagnostic models to analyze large-scale educational assessment data and generate actionable insights that support teaching, learning, and personalized feedback in engineering education. He also examines assessment practices in the era of generative artificial intelligence, emphasizing adaptive, interpretable, and learning-centered systems.

Suman Saha, Pennsylvania State University

Suman Saha received a Ph.D. degree in computer science from Pierre and Marie Curie University, Paris, France, in 2013.

He is an Assistant Teaching Professor at Pennsylvania State University, University Park, PA, USA. Since 2016, he has been involved in academia. Before his current role with Pennsylvania State University, he taught at Illinois Institute of Technology, Chicago, IL, USA, and Illinois State University, Normal, IL, USA.

Dr. Saha received the distinguished William C. Carter Award for his substantial contributions to dependable and secure computing during his doctoral studies. He has held positions at several prestigious National and International Research Labs, including Microsoft, Cambridge, U.K., Harvard University, Cambridge, USA, and the National Institute for Research in Digital Science and Technology (Inria), Paris, France.

Personalization Strategies in Large Language Model-Based Educational Systems: A Scoping Review

Abstract

Traditional education often relies on a one-size-fits-all approach that fails to address diverse student learning needs, motivating the use of personalized learning, defined as the adaptation of instructional content, feedback, and learning pathways to individual learners. Recent advances in large language models (LLMs) enable scalable personalization; however, existing literature lacks a systematic synthesis of how personalization is operationalized across educational contexts, including project-based learning, labs, and large lecture settings. To address this gap, we conduct a scoping review of LLM-based personalization strategies. Using a structured keyword search via the Semantic Scholar API, 366 studies were identified and screened using predefined inclusion and exclusion criteria, resulting in a final corpus of 26 empirical studies. Analysis of these studies reveals diverse applications of LLMs in personalized tutoring, curriculum development, instructional adaptation, learner profiling, prerequisite modeling, and knowledge gap identification, as well as instructor-facing uses such as tailored problem generation and misconception targeting; applications also extend to specialized domains such as robotic surgical training through adaptive visual question-answering frameworks. These findings highlight the potential of LLMs to support scalable, adaptive learning across varied educational settings while augmenting instructional practices. This review contributes a structured synthesis of personalization strategies and identifies key directions for future work, including the need for rigorous evaluation frameworks, improved reliability, and consideration of ethical and implementation challenges in LLM-based educational systems.

Introduction

The rapid advancement of Artificial Intelligence (AI) has led to its adoption across many domains, including finance, engineering, and education [1]. Within educational contexts, AI technologies are increasingly used to support instructional activities such as assessment, feedback generation, and content adaptation [2]. Despite these advances, many instructional environments, particularly large-enrollment and online courses (>100 students), continue to rely on uniform instructional materials and fixed pacing [3]. Such one-size-fits-all approaches often fail to account for differences in students' prior knowledge, learning pace, and cognitive needs, which can lead to disengagement, inequitable learning experiences, and uneven skill development [3].

Recent progress in Large Language models (LLMs) has created new opportunities to address these longstanding challenges. LLMs demonstrate strong capabilities in natural language understanding and generation, enabling them to adapt explanations, feedback, and learning activities in ways that were previously difficult to scale [4]. These capabilities make LLMs particularly well suited for personalized learning scenarios, where instructional content can be dynamically adjusted based on individual student characteristics and performance, including through AI-supported delivery of targeted microlearning resources such as bite-sized videos [28]. At the same time, the growing use of AI in education has reinforced the importance of maintaining instructors as central actors in the learning process rather than replacing them with automated systems. LLM-based systems are most effective when they augment instructional decision-making by providing analytics, targeted feedback, adaptive content, and AI-supported microlearning resources, while instructors continue to provide the pedagogical judgment, emotional support, and contextual guidance that automated systems cannot replicate [3][27]. When designed and deployed appropriately, this human-AI partnership has the potential to improve student engagement, motivation, learning outcomes, and equity across diverse educational settings [5].

Although a growing body of research has explored AI-driven personalization in education, including adaptive tutoring systems and real-time feedback tools, the literature lacks a comprehensive synthesis of how personalization is operationalized in LLM-based educational systems across different contexts [6]. In particular, existing work does not systematically examine the range of personalization strategies employed, the educational settings in which they are deployed, or the trade-offs associated with different technical and pedagogical approaches. This gap limits the ability of educators and researchers to make informed decisions about system design, adoption, and future research directions.

To address this need, we present a scoping review of personalization strategies used in LLM-based educational systems. This review examines how personalization is implemented across diverse educational settings, identifies recurring design patterns and emerging trends, and highlights open challenges related to scalability, accuracy, and instructional integration. By synthesizing current empirical work, this paper aims to support researchers in identifying promising research directions and assist educators in understanding how LLM-based personalization strategies can be aligned with instructional goals and practical constraints. Guided by these objectives, this scoping review is structured around the following research questions:

RQ1. What personalization strategies are used in LLM-based educational systems, and how are these strategies distributed across different educational settings?

RQ2. What benefits, limitations, and trade-offs are associated with different LLM-based personalization strategies in terms of instructional effectiveness, technical complexity, and practical deployment?

Related work

The rapid advancement of AI has generated substantial scholarly interest, resulting in a growing body of research examining the role of AI and LLMs in education. In this section, we synthesize findings from recent systematic reviews and surveys to situate our scoping review within the existing literature and to clarify the specific gap it addresses. We deliberately focus on review-based studies, as they provide a comprehensive view of prevailing research trends, methodological emphases, and unresolved challenges in AI-supported education.

Several recent surveys have examined the broad integration of LLMs in educational contexts. Wang et al. [21] presented an extensive technological survey of LLM applications in education, covering instructional use cases, student-facing support tools, adaptive learning frameworks, datasets, and benchmarks. Their work offers a useful taxonomy of LLM-enabled educational systems and highlights key risks such as bias, privacy concerns, and hallucinations. While this survey provides a thorough overview of technological capabilities, it does not examine how personalization is operationalized at the system level or differentiate among specific personalization strategies.

Similarly, Leigeb et al. [22] analyzed 111 studies published between 2014 and 2024 to examine the benefits and challenges of AI and LLM adoption in higher education. Their review emphasizes outcomes such as adaptive feedback and instructional automation, alongside concerns related to governance, ethics, and explainability. The authors propose institutional pillars for responsible AI integration, with a strong focus on policy and organizational readiness. However, their analysis remains largely conceptual and does not investigate the technical or architectural mechanisms through which personalization is implemented.

Other reviews have examined personalization more directly, but within narrower scopes. Reihanian et al. [23] conducted a scoping review of 32 studies focused on the personalization of computer science education using generative AI. They categorized personalization approaches into intelligent tutoring systems, tailored learning material generation, targeted feedback, AI-augmented assessment, and automated code review. Their findings suggest that systems incorporating explanation-first guidance, structured hinting, delayed solutions, and artifact grounding (anchoring AI responses in specific student work or instructional materials) are more effective than systems that lack these features. While highly relevant, this review is limited to computer science education and does not consider personalization strategies across broader educational contexts or disciplines.

Related work has also explored personalization through the lens of tutoring systems. Liu et al. [24] reviewed 86 studies on intelligent tutoring systems and robotic tutoring systems, examining how AI techniques support cognitive and affective learning processes. Their review distinguishes between systems that emphasize adaptive cognitive support and those designed to provide emotional or social interaction. Although this work situates LLM-based personalization within the broader ITS landscape, LLMs are treated as one of many enabling technologies rather than as the primary focus of personalization design. Maity and Deroy [25] similarly examined the integration of generative AI into modern tutoring systems, outlining strategies such as dynamic content generation, automated problem creation, tailored feedback, and adaptive learning pathways. While their review directly links LLM capabilities to tutoring applications, it does not systematically categorize emerging personalization strategies such as retrieval-augmented generation, knowledge graph–based modeling, or agent-based architectures.

Complementing these perspectives, Moore and Tsay [26] reviewed 26 studies published between 2022 and 2024 that examined AI and LLM use in higher education. Their analysis highlights applications related to personalized feedback, chatbot-based guidance, predictive analytics for identifying at-risk students, and instructional support aimed at improving engagement and motivation. Although this review connects LLM use to student success initiatives, it does not provide a detailed examination of the underlying personalization mechanisms that enable these interventions.

Taken together, existing reviews offer valuable insights into LLM capabilities, intelligent tutoring systems, personalization benefits, and ethical considerations. However, they also reveal a clear gap. Prior work either provides broad overviews of AI in education without differentiating personalization strategies [21, 22], limits analysis to specific disciplines or educational levels [23], or treats LLMs as a component within larger tutoring frameworks rather than as drivers of distinct personalization approaches [24, 25]. To address this gap, our scoping review systematically maps personalization strategies employed in LLM-based educational systems across diverse educational contexts. By focusing on how personalization is implemented, rather than on outcomes or institutional policy alone, this work contributes a needed synthesis that supports both future research and informed educational practice.

Research method

This study employed a scoping review approach to examine recent research on personalization strategies in LLM-based educational systems. This approach was selected to provide a broad overview of how personalization is implemented across studies, rather than to evaluate the effectiveness of a single instructional intervention. The review methodology was intentionally designed to address the research questions by enabling systematic identification, classification, and comparison of personalization strategies across educational contexts.

Literature search strategy

We conducted a two-phase literature search focusing on academic journal articles, conference papers, and preprints. In the first phase, relevant studies were identified using Semantic Scholar, a search engine for scientific papers. We performed a structured keyword-based search using combinations of terms related to personalization and learning strategies, LLM technologies, and educational contexts. The search categories, Boolean connectors, and combined keyword terms are presented in Table 1. This search strategy was designed to capture a wide range of relevant studies while excluding non-educational literature. The initial keyword-based search returned 366 studies. The studies identified through this process were then subjected to a multi-stage screening procedure to ensure relevance and quality.

Table 1. Keyword Search Strings

Category	Connector	Search Terms (Combined with OR)
Personalization and Learning Strategies		adaptive learning | personalized learning | intelligent tutoring | student modeling | learner model | knowledge tracing | curriculum sequencing | learning path | adaptive

		feedback | recommendation
LLM Technologies	+ (AND)	large language model | LLM | ChatGPT | GPT-4 | foundation models
Educational Contexts	+ (AND)	education | classroom | higher education | K-12 | student | teacher | pedagogy
Excluded Terms	− (NOT)	survey | reviews | systematics review | scoping review

Screening and study selection

All retrieved studies were imported into Rayyan, an online tool for managing systematic and scoping reviews. Title and abstract screening were conducted using predefined inclusion and exclusion criteria related to research focus, publication year, educational setting, publication venue, and language, as summarized in Table 2. Consistent with the objective of capturing recent advances in AI-driven educational personalization, the review was restricted to studies published in 2024 and 2025. The screening process focused on studies that implemented AI techniques for personalizing education, with attention to applications in engineering education contexts. Studies categorized as surveys, systematic reviews, or scoping reviews were excluded.

Following title and abstract screening, a full-text review was conducted to further assess study relevance and alignment with the objectives of this review. This step ensured that selected studies included concrete implementations of AI-based personalization in educational settings. Application of these criteria resulted in a final corpus of 26 studies. The study selection process is summarized in Figure 1, which illustrates the number of records identified, screened, excluded, and included at each stage of the review.

Table 2. Inclusion/Exclusion criteria

Focus	AI in education for personalized learning
Time of Publication	2024 and 2025
Educational Setting	Higher Education, Online Courses, K-12, and Specialized Training Programs
Publication Venue	Peer reviewed articles, conference papers, journal articles, and preprints
Language	English

Data extraction and analysis

Relevant information from the 26 selected studies was extracted using a structured data charting process. To support systematic organization, we developed a customized charting form in Microsoft Excel that served as a standardized template for mapping key elements from each study. The charting form included publication details, educational context, personalization strategies employed, reported outcomes, and any emerging trends or gaps identified by the authors. To ensure consistency in categorization, each study was assigned a primary personalization strategy based on the dominant mechanism driving personalization within the system. When multiple strategies were present, the dominant strategy was selected to represent the study. This charting approach facilitated thematic analysis and supported consistent categorization of findings across the three dimensions presented in the Results section: publication venues, educational settings, and personalization strategies.

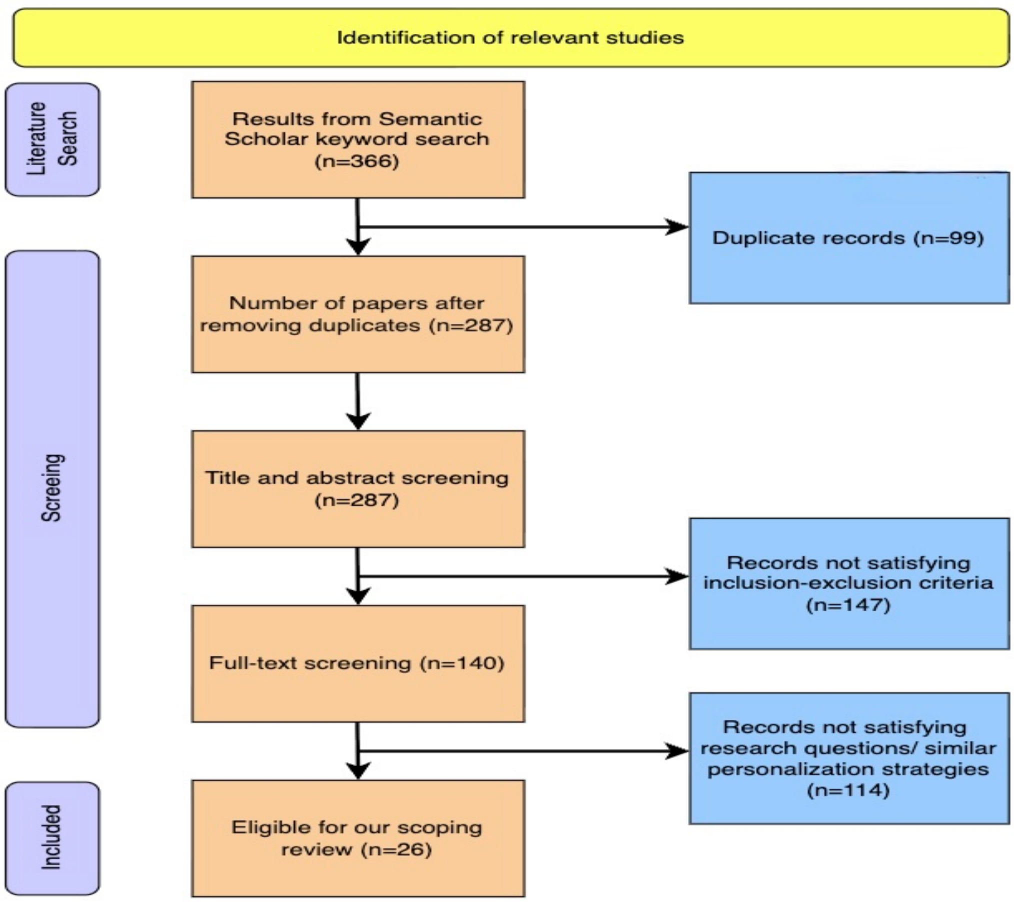

Figure 1: PRISMA-style flow diagram illustrating the study selection process from 366 identified records to 26 included studies after duplicate removal, screening, and full-text assessment.

Results

This section presents the results of the 26 empirical studies identified through the scoping review. The results are organized along three analytical dimensions that directly address the research questions. Specifically, the analysis of publication venues and educational settings provides context for RQ1, which examines how personalization strategies are distributed across

LLM-based educational systems and instructional environments. The analysis of personalization strategies identifies the technical mechanisms used to support personalization and provides the empirical basis for RQ2, which examines the benefits, limitations, and trade-offs associated with these approaches. The findings reported in this section are descriptive in nature and establish the foundation for the comparative interpretation presented in the Discussion section.

Publication venue

Table 3 summarizes the distribution of the 26 selected studies across publication venues. The studies appear in a diverse set of outlets, including education-focused conferences and journals, AI venues, and interdisciplinary publication forums. In addition to the diversity of publication venues, the selected studies also exhibit a broad geographical distribution, further underscoring the global and interdisciplinary nature of research on LLM-based personalization. The corpus includes contributions from several individual countries, including the United States, China, the United Kingdom, Japan, Iran, India, Singapore, and the United Arab Emirates. Moreover, a number of studies reflect international collaboration, with co-authored works spanning multiple regions, such as collaborations between Finland, Hong Kong, and Switzerland, as well as between the United States and China. This distribution reflects the cross-disciplinary nature of research on LLM-based personalization, where advances are informed by both educational research and developments in AI systems.

The presence of studies across multiple venue types suggests that LLM-based educational personalization is being investigated from different perspectives, including pedagogical design, system development, and applied evaluation. At the same time, the inclusion of preprints alongside conference and journal publications indicates that this area is still evolving, with many systems and approaches being reported during early stages of dissemination. Importantly, the venue distribution reflects research activity and scholarly interest rather than the extent of classroom adoption or instructional impact, which are examined more cautiously in later sections.

Table 3. Publication Venues

Publication Venue	Number of studies
IEEE International Conference on Future Machine Learning and Data Science	1
IEEE International Conference on Quantum Computing and Engineering	1
Associate for Computing Machinery (ACM)	5
IEEE Transaction on Medical Imaging	1
International Journal of Scientific Research and Management	1

IEEE Region 10 Symposium	1
Association for Computational Linguistics	2
International Conference on AI in Education	1
International Conference on Computational Linguistics	1
IEEE International Conference on Educational Technology	1
Learning at Scale	1
IEEE Conference of AI	1
AAAI Conference of AI	1
OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 5.0	1
Preprints	7

Educational setting

The 26 selected studies span multiple educational settings, as summarized in Figure 2. Most implementations of LLM-based personalized learning systems are situated in higher education contexts (n = 15), followed by K-12 settings (n = 6), Massive Open Online Courses (MOOCs) (n = 3), and specialized training programs (n = 2). Figure 2 visualizes this distribution and highlights the concentration of current research efforts in postsecondary environments.
The predominance of higher education settings suggests that universities and colleges provide favorable conditions for experimenting with LLM-based personalization, including greater instructional autonomy, access to technical infrastructure, and fewer regulatory constraints compared to K-12 environments. In contrast, the smaller number of studies in K–12 and MOOC settings may reflect challenges related to data privacy, age-appropriate deployment, scalability, and curriculum standardization. The presence of studies in specialized training programs indicates emerging interest in applying LLM-based personalization beyond traditional academic contexts, particularly for targeted skill development.

Personalization strategies

Analysis of the 26 empirical studies identified five recurring personalization strategies used in LLM-based educational systems: Prompting, Retrieval-Augmented Generation (RAG), Knowledge Graphs (KGs), Agent-based Architecture, and Dynamic Adaptation. These strategies represent distinct design mechanisms through which personalization is implemented, rather than differences in instructional outcomes. Figure 3 illustrates the percentage distribution of these strategies across the reviewed studies, showing that Prompting and RAG are the most frequently reported approaches (n = 6 each), followed by Knowledge Graphs (n = 5), Agent-based Architectures (n = 5), and Dynamic Adaptation approaches (n = 4). The distribution depicted in Figure 3 reflects variation in system complexity, data requirements, and levels of automation. The following subsections describe each personalization strategy in detail, highlighting their implementation characteristics, potential benefits, and associated challenges.

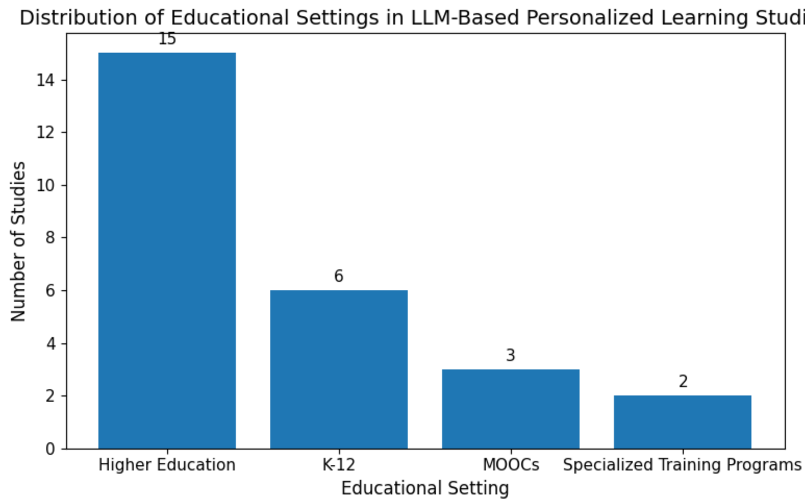

Figure 2: Distribution of educational settings represented in the 26 studies included in the scoping review. Most implementations of LLM-based personalized learning systems are situated in higher education, with fewer studies conducted in K-12, MOOC, and specialized training contexts.

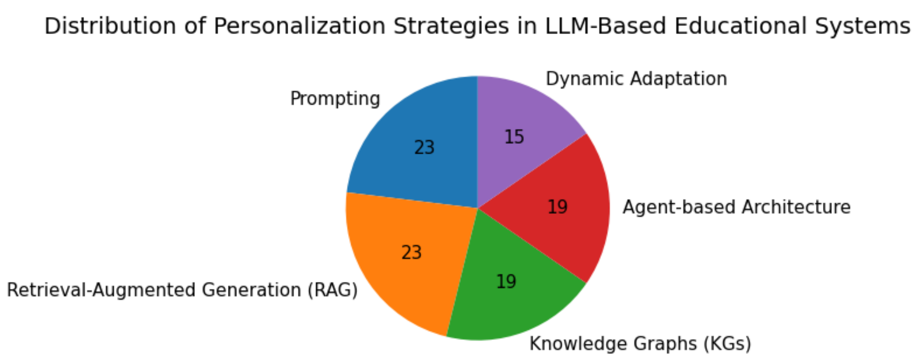

Figure 3: Percentage distribution of personalization strategies reported in the 26 LLM-based educational systems reviewed, illustrating the relative prevalence of prompting, retrieval-augmented generation, knowledge graph–based, agent-based, and dynamic adaptation approaches.

Prompting: Prompting refers to the practice of composing natural language instructions to guide LLMs in generating task-specific outputs without modifying model parameters. From an educational perspective, prompting is particularly attractive because it requires minimal technical overhead and allows instructors to directly shape model behavior using pedagogically meaningful instructions, such as learning objectives, rubrics, and difficulty constraints.

Technically, prompting is also practical because the large parameter space of LLMs makes it infeasible for humans to identify and adjust optimal parameters for specific natural language processing tasks. As a result, extensive research over recent years has focused on developing effective prompting techniques [7].

One prominent prompting technique identified in the reviewed literature is Zero-Shot Chain-of-Thought prompting, in which LLMs are instructed to reason step by step to decompose complex problems into smaller, human-like reasoning steps. This approach has been applied in the context of grading Massive Open Online Courses (MOOCs), where LLMs were used to automate scoring and feedback. In these systems, models generated step-by-step reasoning to assign grades and provide justifications for score deductions based on instructor-provided rubrics, while also offering constructive feedback aligned with identified student errors [8].

Another prompting technique observed in the literature is guided prompting, which emphasizes close collaboration between instructors and LLMs. In this approach, instructors supported the generation of introductory coding problems, feedback, and misconception tags by supplying structured prompt templates. These templates included predefined elements such as learning outcomes, student difficulty levels, target audience, and contextual factors, enabling instructors to retain control over pedagogical intent while leveraging the generative capabilities of LLMs [9]. Across both instructor-facing tools, grading accuracy and problem curation performance were reported to be comparable to, or better than, traditional approaches when used in conjunction with human oversight [8, 9]. These findings suggest that prompting can function as a viable personalization strategy that augments instructional practice rather than replacing instructor involvement.

Beyond instructor-facing applications, prompt-based educational systems have also been used to directly support student learning. Alipour et al. [20] employed a student profiling approach in which anonymized learner attributes, including age, prior programming experience, and cognitive indicators, were incorporated into a three-part prompt structure consisting of explanation, example, and exercise components. The explanation and example components were used to control response difficulty and illustrative content, while the exercise component provided scaffolded guidance toward problem solutions. This system was deployed in a CS1 introductory programming course, where results indicated improvements in student engagement, confidence, and quiz performance [20]. At the same time, the study highlighted risks related to over-reliance on AI-generated assistance and potential hallucinations, reinforcing the need for continued human oversight in prompt-based educational systems.

Retrieval-Augmented Generation (RAG): Retrieval-Augmented Generation (RAG) is a technique that extends the capabilities of LLMs by combining parametric knowledge encoded within model parameters with non-parametric knowledge retrieved from external sources, such as databases, PDFs, and instructional materials. By grounding generation in externally retrieved content, RAG enables LLM-based systems to produce responses that are more up to date and contextually relevant [10].

Among the reviewed studies, RAG was implemented in the development of *Study Pilot*, an AI-powered personalized learning platform. In this system, ChromaDB, a vector database, was used to store vector embeddings derived from external resources, including databases, PDF documents, and YouTube videos, forming a structured knowledge base. A similarity search algorithm was then applied to compute relevance scores between a learner's natural language query and stored content, enabling the system to generate personalized responses across multiple formats, such as conversational explanations, quizzes, and podcasts, to accommodate diverse learning preferences. Empirical results reported in the study indicate that RAG-based knowledge extraction and response generation achieved higher relevance and accuracy compared to traditional LLM-based educational systems that rely solely on parametric knowledge [11].

At the same time, these accuracy gains come with increased infrastructure and maintenance requirements, as RAG-based systems depend on curated external data sources, embedding pipelines, and retrieval mechanisms, which introduce additional institutional investment beyond standalone LLM deployments.

Knowledge graphs (KGs): Knowledge Graphs provide a structured representation of data by modeling entities and their relationships in the form of a graph, where nodes represent entities and edges capture interactions between them. This graph-based representation is particularly useful for efficient querying, semantic search, and reasoning over structured educational content [12]. Among the reviewed studies, knowledge graphs were constructed from a K-12 algebra textbook to model topics, subtopics, and prerequisite relationships as interconnected nodes and edges. This structure enabled an LLM-based system to analyze prior student interactions in order to estimate learners' knowledge states and rank prerequisite questions accordingly.

The derived knowledge state information and prerequisite question rankings were then appended to prompts to support tiered guidance, providing advanced assistance to proficient students and more detailed explanations for students requiring additional support [13]. In the reported experiment, which involved six participants, feedback quality varied across problem difficulty levels. While feedback aligned well with provided solutions for easy problems, performance deteriorated for medium and hard problems, indicating limitations in the system's ability to generalize across increasing task complexity. Because these findings are based on a small sample size, the results should be interpreted cautiously, as they limit the extent to which conclusions about the effectiveness and robustness of knowledge graph–based personalization can be generalized beyond the studied context [13].

Agent-based architecture: Agent-based architectures are designed to address educational scenarios that require sustained reasoning, adaptive decision-making, and multiple forms of instructional support that cannot be effectively handled by a single model or static workflow. Agent-based architecture consists of one or more LLM-powered agents, where each agent is assigned a specialized role and interacts with other agents to support coordinated decision-making. This design offers a more reasoning-enhanced framework, as single-agent systems often struggle with complex, multi-stage educational tasks [14].

Several reviewed studies applied multi-agent systems within educational contexts. One study proposed a multi-agent neuro-symbolic AI framework for K-12 biology education and university-level digital game-based learning. The system comprised three agents: a Tutor Agent, a Peer Agent, and an Educational Ontology. The Educational Ontology functioned as a knowledge graph representing domain-specific concepts, difficulty levels, pedagogical rules, and their relationships. The Tutor Agent employed a reinforcement learning mechanism to dynamically regulate the level of instructional support based on a student's demonstrated proficiency on specific concepts. In contrast, the Peer Agent focused on providing emotional support and tailoring responses based on the learner's emotional state, simulating peer collaboration. Together, these agents worked collaboratively to deliver a personalized learning experience that integrated cognitive and affective dimensions of learning [15].

Similarly, another study introduced *Agent4Edu*, an agent-based system that integrated learner profiles with a memory module to support personalization over time. The learner profile captured relatively stable learning preferences as well as cognitive factors that evolved with practice and experience. The memory module consisted of short-term and long-term components. Short-term memory retained recent interactions to support coherent, context-aware decision-making, while long-term memory stored reinforced knowledge, high-level reflective summaries of learning progress, and dynamic assessments of concept mastery. A forgetting mechanism based on decay functions accounted for time and practice frequency, allowing the system to model knowledge degradation realistically. These memory components interacted with learner profiles and action modules, enabling LLM-powered agents to retrieve past experiences, update internal states, and generate adaptive responses tailored to personalized learning algorithms in intelligent educational systems [16].

Dynamic adaptation: Dynamic adaptation refers to the real-time personalization and sequencing of learning activities based on individual student performance and preferences [17]. Among the reviewed studies, we identified the Dynamic Skill Adaptation (DSA) framework, which adopts a human-inspired personalization strategy by decomposing complex skills into a dependency-based skill graph. These skill graphs organize topics and their prerequisites hierarchically and are used to generate textbook-like instructional descriptions. The model is first trained on these descriptions in increasing order of difficulty and subsequently trained on exercise-style data, allowing model parameters to be adjusted to generate educational content that adapts to real-time student performance [18].

The evaluation of DSA focused primarily on system-level performance rather than direct measures of student learning outcomes. Specifically, DSA was implemented across multiple open-source LLMs and compared against baseline models. Both DSA-enhanced and baseline models were evaluated using standardized datasets under different experimental conditions, with results showing that DSA-based models achieved significantly higher performance scores than their baseline counterparts [18]. These results demonstrate improvements in model behavior and content generation quality, but they do not directly measure changes in student learning or instructional effectiveness.

In contrast, another study proposed the Personalized AI-Powered Progressive Learning Platform (PAPPL), which emphasizes learner-level adaptation through continuous monitoring of student interactions within an integrated Intelligent Tutoring System (ITS). PAPPL leverages GPT-4o as the core tutor module to analyze multiple student attempts on web-based problems and detect recurring misconceptions using a learner-state analyzer that maintains temporally updated profiles of student progress and weaknesses. The system generates context-sensitive, question-based hints that progressively increase in specificity, beginning with broad prompts to encourage metacognitive reasoning and becoming more targeted as challenges persist. This adaptive behavior is supported through multi-layered prompt engineering that incorporates problem context, historical interactions, instructor explanations, and related content, while a content manager and analytics dashboard provide instructors with evidence-based insights into class-wide learning trends [19].

PAPPL was evaluated with student participants from the University of Illinois Urbana–Champaign (UIUC) and George Washington University (GWU) using a set of 25 pavement engineering questions, a topic for which participants had no prior background knowledge. Experimental results indicated that students who received feedback from PAPPL demonstrated stronger cognitive performance and reported higher levels of engagement, accuracy, and satisfaction compared to comparison conditions [19]. Together, these studies illustrate that dynamic adaptation approaches can be evaluated at different levels, ranging from system-level performance gains to learner-level educational outcomes.

Discussion

This scoping review examined how personalization is realized in LLM-based educational systems and what trade-offs accompany different implementation strategies. By synthesizing findings across 26 recent empirical studies, the review provides insight into both the distribution of personalization strategies across educational contexts and the practical implications of adopting these approaches in instructional settings.

Addressing RQ1: Distribution of personalization strategies across contexts

Findings related to publication venues and educational settings indicate that research on LLM-based personalization is primarily concentrated in higher education, with more limited exploration in K-12, MOOCs, and specialized training contexts. This distribution reflects both opportunity and constraint. Higher education environments often provide greater flexibility for experimentation, access to technical infrastructure, and fewer regulatory barriers, which may explain their dominance in current implementations. At the same time, the presence of studies in K-12 and nontraditional learning environments suggests growing interest in extending LLM-based personalization beyond postsecondary education, albeit at a slower pace.

Across these contexts, the relatively even distribution of personalization strategies identified in the Results section suggests that the field has not converged on a single dominant approach. Instead, researchers appear to be exploring multiple design pathways in parallel. This diversity indicates that personalization in LLM-based educational systems is highly context-dependent, shaped by instructional goals, learner populations, and institutional constraints rather than by a universally optimal technical solution.

Addressing RQ2: Trade-offs among personalization strategies

Analysis of personalization strategies highlights clear trade-offs between accessibility, accuracy, system complexity, and resource requirements. Prompting-based approaches emerge as the most accessible strategy, requiring minimal infrastructure beyond access to an LLM and allowing instructors to retain direct pedagogical control through prompt design [8, 9]. These characteristics make prompting particularly attractive for resource-constrained settings and for early-stage adoption. However, reliance on parametric knowledge introduces risks related to factual inaccuracies and hallucinations, reinforcing the importance of instructor oversight when prompting is used to support learning activities.

Retrieval-Augmented Generation addresses these limitations by grounding model outputs in external knowledge sources, improving relevance and accuracy in personalized responses [10, 11]. From an educational perspective, RAG is especially valuable in domains where content evolves rapidly or where precise alignment with curricular materials is required. At the same time, the benefits of RAG come with increased institutional investment, including the need for curated content repositories, embedding pipelines, and ongoing maintenance. These requirements may limit feasibility in settings lacking technical support or content management capacity.

Knowledge graph–based systems and agent-based architectures represent structured approaches to personalization. By explicitly modeling domain knowledge, learner states, or instructional roles, these strategies enable deeper reasoning about learning progress and instructional sequencing [13, 15, 16]. However, the reviewed studies reveal important limitations. Knowledge graph approaches often rely on small-scale evaluations and require extensive domain modeling, raising concerns about scalability and generalizability. Similarly, agent-based systems introduce interpretability challenges as decision-making becomes distributed across interacting components, complicating instructor oversight and pedagogical validation.

Dynamic adaptation strategies extend personalization further by continuously adjusting instruction based on learner performance over time [17, 19]. Importantly, the reviewed studies illustrate that evaluation practices vary substantially within this category. Some approaches focus on system-level performance gains, while others assess learner-level outcomes such as engagement and cognitive growth. This distinction underscores the need for clearer evaluation frameworks that align technical improvements with educational impact. While dynamic adaptation shows strong promise, existing evidence remains limited to specific domains, highlighting the need for broader validation across disciplines and learner populations.

Implications for research and practice

Taken together, the findings suggest that no single personalization strategy is universally applicable. Instead, effective deployment of LLM-based personalization requires alignment among instructional goals, available resources, and desired levels of automation. For educators and institutions, this underscores the importance of viewing LLM-based personalization as a design space rather than a fixed solution. Human oversight remains a central consideration across all strategies, both to ensure pedagogical soundness and to mitigate risks related to bias, error propagation, and over-reliance on automated systems.

Overall, this scoping review positions LLM-based personalization as a rapidly developing area with significant potential to shape future educational practice. By articulating the trade-offs associated with existing strategies and identifying gaps in current evidence, this work provides a foundation for more systematic, context-aware research on personalized learning with LLMs.

Future work

This scoping review highlights several directions for future research that are essential for advancing LLM-based personalized learning in engineering and computing education. First, while this review categorizes existing personalization strategies, there is a clear need for systematic comparative studies that evaluate these approaches within comparable instructional contexts. Such studies would help clarify relative strengths and limitations, as well as identify when lightweight strategies, such as prompting, are sufficient and when more complex approaches, such as retrieval-augmented or agent-based systems, are pedagogically justified.

Second, future work should place greater emphasis on evaluation frameworks that explicitly distinguish between system-level performance improvements and learner-level educational outcomes. Emerging work on GenAI-supported learning environments also suggests the importance of measuring outcomes such as learner engagement, confidence, reflection, and feedback quality in authentic instructional settings [29]. Many existing studies report gains in model behavior or response quality without directly measuring learning impact, making it difficult to assess educational effectiveness across contexts. Larger-scale evaluations, particularly in K-12 and informal learning environments, are also needed to assess generalizability beyond higher education settings.

Finally, ethical and practical considerations warrant sustained attention. As personalization systems increasingly rely on learner data and adaptive decision-making, future research should address transparency, bias, learner agency, and instructor oversight. In addition, greater attention to interpretability and instructor-facing explanations may improve trust and support adoption of more complex systems, such as agent-based and graph-driven architectures. Developing practical design guidelines and implementation frameworks that help educators select and deploy personalization strategies based on instructional context, resource availability, and pedagogical intent will be critical for broader real-world adoption.

Conclusion

This paper presented a scoping review of personalization strategies used in LLM-based educational systems, synthesizing findings from 26 recent empirical studies across publication venues, educational settings, and system designs. The review demonstrates that personalization in LLM-based systems is characterized by methodological diversity rather than convergence on a single dominant approach. Instead, different strategies reflect distinct trade-offs among accessibility, accuracy, system complexity, and resource requirements.

For engineering and computing education, these findings emphasize the importance of context-aware personalization grounded in pedagogical goals rather than purely technical considerations. Across strategies, human oversight consistently emerges as a central design principle, reinforcing the role of instructors as active partners in AI-supported learning environments. Rather than replacing instructional judgment, current LLM-based personalization approaches are most effective when they augment educators' ability to manage scale, provide timely feedback, and support diverse learners.

By clarifying practices and identifying trade-offs, this scoping review provides a foundation for more systematic and pedagogically grounded research on personalized learning with LLMs. As LLM technologies continue to evolve, aligning technical innovation with educational intent and empirical evidence is critical to realizing their potential in engineering education.

References

[1] Z. Bahroun, C. Anane, V. Ahmed, and A. Zacca, "Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis," *Sustainability*, vol. 15, no. 17, Art. no. 12983, 2023, doi: 10.3390/su151712983

[2] C. Liu, "A comprehensive review of applications of AI technologies in higher engineering education," *Discover Education*, vol. 4, Art. no. 528, 2025, doi: 10.1007/s44217-025-00954-0

[3] Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promise and implications for teaching and learning*. Center for Curriculum Redesign.

[4] M. Sharples, Y. H. Teo, S. K. Goh, T. W. Chan, and C. Yin, "Large language models: A survey," *arXiv preprint*, arXiv:2402.06196, 2024

[5] T. Alqahtani, H. Badreldin, M. Alrashed, T. Alqadi, S. Alshammari, S. Alshammari, M. Alshammari, and W. Albaker, "The impact of artificial intelligence (AI) on students' academic performance: A systematic review," *Education Sciences*, vol. 15, no. 3, Art. no. 343, 2025, doi: 10.3390/educsci15030343

[6] Y. Shi, K. Yu, Y. Dong, and F. Chen, "Large language models in education: A systematic review of empirical applications, benefits, and challenges," *Computers and Education: Artificial Intelligence*, vol. 10, 2025, doi: 10.1016/j.caeai.2025.100016

[7] S. Vatsal and H. Dubey, "A survey of prompt engineering methods in large language models for different NLP tasks," *arXiv preprint*, arXiv:2407.12994, 2024

[8] S. Golchin, N. Garuda, C. Impey, and M. Wenger, "Grading massive open online courses using large language models," *arXiv preprint*, arXiv:2406.11102, 2024

[9] M. Hoq, J. Vandenberg, S. Jiao, S. Lee, B. Mott, N. Norouzi, J. Lester, and B. Akram, "Facilitating instructors–LLM collaboration for problem design in introductory programming classrooms," *arXiv preprint*, arXiv:2504.01259, 2025

[10] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," arXiv preprint, arXiv:2012.09313, 2020

[11] R. T. Karnati, H. Kundra, and P. R. S. S. Naidu, "Study Pilot: An AI-powered platform for personalized learning through retrieval-augmented generation on diverse user content," International Journal of Scientific Research in Engineering and Management, vol. 9, no. 1, pp. 1–6, 2025

[12] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. De Melo, C. Gutierrez, S. Kirrane, J. E. Labra Gayo, R. Navigli, Y. Nechaev, A. C. Ngonga Ngomo, A. Polleres, E. Ruckhaus, J. Sequeda, S. Staab, and A. Zimmermann, "Knowledge graphs," *ACM Computing Surveys*, vol. 54, no. 4, pp. 1–37, 2021, doi: 10.1145/3447772

[13] P. Ocheja, B. Flanagan, Y. Dai, and H. Ogata, "How good is ChatGPT in giving adaptive guidance using knowledge graphs in e-learning environments?" *arXiv preprint*, arXiv:2412.03856, 2024

[14] H. Huang, Y. Zhang, H. Liu, S. Chen, and J. Zhu, "A survey on LLM-based multi-agent systems: Recent advances and new frontiers in application," *arXiv preprint*, arXiv:2412.17481, 2024

[15] J. Hare and Y. Tang, "Toward generalized autonomous agents: A neuro-symbolic AI framework for integrating social and technical support in education," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025

[16] L. Gao, Y. Zhang, X. Liu, Z. Li, and H. Wang, "Agent4Edu: Generating learner response data by generative agents for intelligent education systems," in *Proceedings of the ACM Conference on Learning @ Scale*, 2025

[17] F. Jurado, O. C. Santos, M. A. Redondo, J. G. Boticario, and M. Ortega, "Providing dynamic instructional adaptation in programming learning," in *Hybrid Artificial Intelligence Systems*, Lecture Notes in Computer Science, vol. 5271, E. Corchado, A. Abraham, and W. Pedrycz, Eds. Berlin, Germany: Springer, 2008, pp. 329–336, doi: 10.1007/978-3-540-87656-4_41

[18] J. Chen and D. Yang, “Dynamic skill adaptation for large language models,” *arXiv preprint*, arXiv:2412.19361, 2024

[19] S. Bafandkar, S. Chung, H. Khosravian, and A. Talebpour, “PAPPL: Personalized AI-powered progressive learning platform,” arXiv preprint, arXiv:2508.14109, 2025

[20] M. Abolnejadian, S. Alipour, and K. Taeb, “Leveraging ChatGPT for adaptive learning through personalized prompt-based instruction: A CS1 education case study,” in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA ’24)*, Honolulu, HI, USA, May 2024, pp. 1–8, doi: 10.1145/3613905.3637148

[21] S. Wang, T. Xu, H. Li, C. Zhang, and J. Liang, “Large language models for education: A survey and outlook,” arXiv preprint, arXiv:2403.18105, 2024, doi: 10.48550/arxiv.2403.18105

[22] T. Leitgeb and M. Leitgeb, “Artificial intelligence and large language models in higher education: Results of a systematic review,” *Ubiquity Proceedings*, Art. no. 201, 2025, doi: 10.5334/uproc.201

[23] I. Reihanian, Y. Hou, and Q. Sun, “From pilots to practices: A scoping review of GenAI-enabled personalization in computer science education,” *AI*, vol. 7, no. 1, Art. no. 6, 2025, doi: 10.3390/ai7010006

[24] V. Liu, E. Latif, and X. Zhai, “Advancing education through tutoring systems: A systematic literature review,” *arXiv preprint*, arXiv:2503.09748, 2025

[25] S. Maity and A. Deroy, “Generative AI and its impact on personalized intelligent tutoring systems,” *arXiv preprint*, arXiv:2410.10650, 2024

[26] W. Moore and L.-S. Tsay, “From data to decisions: Leveraging AI for proactive education strategies,” in *Proceedings of the International Conference on AI Research*, vol. 4, no. 1, 2024, doi: 10.34190/icair.4.1.3082

[27] S. Saha, F. Rahbari, F. Sadique, S. K. C. Velamakanni, M. Farooque, and W. J. Rothwell, “Next-Gen Education: Enhancing AI for Microlearning,” in *Proceedings of the 2025 ASEE Annual Conference & Exposition*, 2025, doi: 10.18260/1-2--56998.

[28] S. K. C. Velamakanni and S. Saha, “Snackable Study: Boosting Micro-learning with Bite-Size Videos,” in *Proceedings of the 2025 ASEE Annual Conference & Exposition*, 2025, doi: 10.18260/1-2--57656

[29] W. Petula, A. Joshi, P. Tu, A. Somasundar, and S. Saha, “DeliverC: Teaching Pointers through GenAI-Powered Game-Based Learning,” in *SIGCSE TS 2026: Proceedings of the 57th*

ACM Technical Symposium on Computer Science Education, pp. 838–844, 2026, doi: 10.1145/3770762.3772509