Discover Sustainability

Review

The role of large language models in personalized learning: a systematic review of educational impact

Sahil Sharma[1] · Puneet Mittal[3] · Mukesh Kumar[2] · Vivek Bhardwaj[1]

Received: 25 December 2024 / Accepted: 25 March 2025
Published online: 04 April 2025

Abstract
The rapid evolution of technology has significantly transformed the educational landscape, with the advent of Large Language Models (LLMs) introducing new possibilities for personalized learning. This systematic review examines the educational impact of LLM-based learning systems compared to traditional educational approaches, focusing on six critical research questions. These questions explore the effectiveness of LLMs in enhancing student engagement, emotional and social development, real-time progress monitoring, and their role in creating fair and rigorous examination environments. Furthermore, the review addresses challenges such as ethical considerations, privacy concerns, and the extent to which LLMs can simulate real-world teaching experiences. A total of 55 studies, published between 2020 and 2024, were systematically analyzed to explore the impact of Large Language Models (LLMs) on educational outcomes, including emotional, social, and academic development. These studies included a combination of peer-reviewed articles, conference papers, and journal publications, which were selected through a set of predetermined inclusion and exclusion criteria. Quality assessment criteria ensured the inclusion of high-quality research focusing on the application of LLM-based AI technology in education. The review also highlights key challenges and limitations, including issues of accessibility, ethical dilemmas, and the integration of AI into traditional education systems. Findings underscore the potential of LLMs to revolutionize education through personalized learning, while also addressing the critical need for rigorous evaluation and ethical deployment to ensure equitable and effective outcomes.

Keywords Large language models (LLMs) · Personalized learning · Educational technology · Student engagement · Artificial intelligence · AI in education

1 Introduction

Education has undergone a remarkable transformation over the centuries, transitioning from oral and handwritten texts to the sophisticated, technology-infused learning environments of today. In the pre-digital era, students relied on face-to-face interactions with educators and printed materials, fostering a classroom-centered model of learning that emphasized rote memorization and repetitive practice [1, 2]. While effective for foundational knowledge, this approach often struggled to address individual student needs and adapt to diverse learning styles [3]. The introduction of computers in education in the late twentieth century marked a pivotal shift. Educational software, interactive tutorials, and early learning management systems (LMS) allowed educators to expand their pedagogical toolkit. Computers became tools

✉ Vivek Bhardwaj, vivek.bhardwaj@jaipur.manipal.edu | [1]School of Computer Science & Engineering, Manipal University Jaipur, Jaipur, India 303007. [2]Advanced Centre of Research & Innovation (ACRI), Department of Computer Application, Chandigarh School of Business, Chandigarh Group of Colleges Jhanjeri, Mohali, Punjab, India 140307. [3]ASET, Amity University Punjab, Mohali, India.

for both teachers and students, enabling access to vast digital repositories of information and laying the groundwork for self-directed learning [4, 5]. Technologies such as video-based tutorials and online discussion forums further enhanced collaborative and experiential learning. However, these systems lacked the adaptability and personalization that AI would later bring to education [6]. The rise of Artificial Intelligence (AI) introduced unpredictable possibilities for new ways of learning. Systems driven by AI could analyze student behavior, adapt content delivery, and provide personalized feedback [7, 8]. Among the most transformative innovations in this domain were large language models (LLMs) like ChatGPT, which emerged as versatile tools capable of simulating human-like conversation and generating high-quality content [7, 9]. By integrating AI into education, students and educators gained access to real-time problem-solving, creative content generation, and enhanced language learning [5, 6]. Nevertheless, concerns about dependency, ethical issues, and equity in access remained prominent [4, 10].

In recent years, the field has witnessed another significant milestone: the development of local, open sourced, online, hybridized architecture pretrained model's, one such architecture is on-device LLMs such as Meta's, Microsoft's, Mistral AI large language models (like Llama 3 and other Companies models: Phi-3, Mistral 7B, Falcon 180B, etc.) [5, 6]. Unlike cloud-based solutions, these models can be operated directly on personal devices, ensuring improved data privacy models for faster response times, and the ability to function offline.

Figure 1 illustrates this progression—from traditional learning methods to cutting-edge AI applications highlighting how these technological advancements are bridging global innovations with localized applications. Local LLMs are particularly significant in addressing educational disparities by enabling resource-limited regions to benefit from AI-driven learning tools [4, 5]. Furthermore, these models empower educators to customize curriculum based on cultural and contextual needs, fostering inclusivity and innovation [5, 10]. The integration of local LLMs into new smart classrooms and personal devices signals a new era of enhanced knowledge acquisition and collaboration. These tools not only support critical thinking and problem-solving but also enable students to actively participate in interdisciplinary projects, nurturing a generation of learners equipped for a rapidly evolving world [4].

As illustrated in Fig. 1, the convergence of technological advancements highlights the transformative potential of innovation in shaping equitable, dynamic, and impactful educational ecosystems. This evolution necessitates collaboration between educators and policymakers to leverage AI responsibly while addressing challenges, ensuring that education remains a powerful tool for universal empowerment and progress [1, 4]. Human ingenuity, with its ability to think outside the box and develop novel approaches, is pivotal. Instead of merely enhancing existing methods, such as making horses run faster, why not invent something entirely new—like a car or another faster mode of transportation? Such groundbreaking ideas have driven rapid technological advancements across various sectors.

Fast forward to today, where a student interacts with a smart, AI-enabled learning environment powered by large language models (LLMs), receiving personalized feedback and adaptive content tailored to their unique learning style. This remarkable shift underscores the revolutionary journey of education, propelled by technological progress and our deepening understanding of how humans learn. Consider the following compelling statistics: A meta-analysis of 50 studies revealed that students using intelligent tutoring systems outperformed 75% of their peers in traditional classrooms [11]. The global AI in education market is projected to reach $20.2 billion by 2027, with a compound

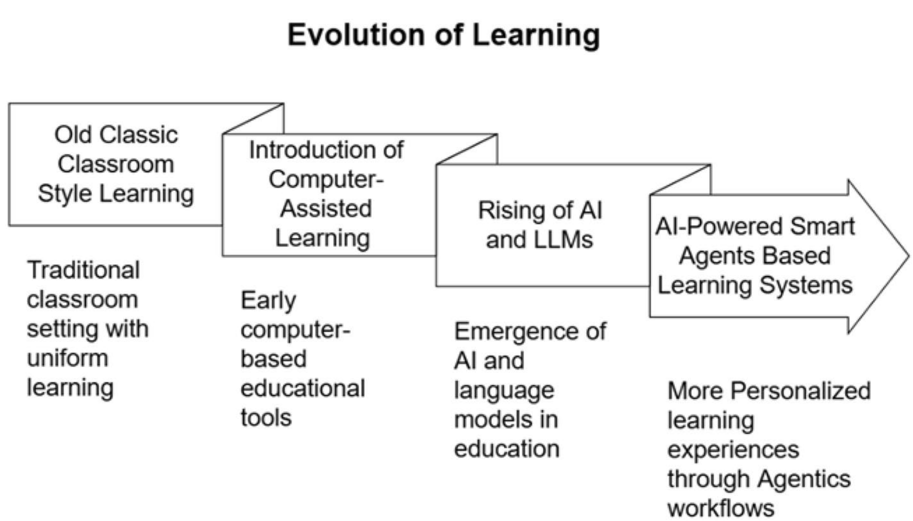


Fig. 1 Evolution of technology in educational history

annual growth rate (CAGR) of 45.12% from 2020 to 2027 [12]. Within the next three years, 47% of learning management tools are expected to integrate AI capabilities [13].

These data points depict a future where AI becomes a fundamental component of education, rather than just an auxiliary tool. Advanced Personalized AI Learning Systems (APALS), including intelligent tutoring systems (ITSs), individualized learning platforms (ILPs), automated grading systems (AGS), and dynamic content generation (DCG) tools, are at the forefront of this revolution. Envision a world where every student has access to a personalized AI tutor—available 24/7, adapting in real time to individual needs and learning speeds. This vision is not science fiction but an imminent reality. To explore this paradigm shift, it is imperative to examine how AI-powered, customized learning systems—particularly those driven by LLMs—enhance student knowledge acquisition compared to traditional teaching methodologies. Understanding their impact on information retention and cognitive development is essential for the future of education, especially as AI-based platforms offer adaptive learning paths, real-time feedback, and tailored instructional strategies.

2 Literatur review

The integration of AI and large language models (LLMs) into the educational field can reshape traditional learning methods, enhancing user-specific personalization and engagement. Many studies reveal that AI-driven systems can adapt to individual student needs, improving academic outcomes through customized content and real-time feedback. Unlike traditional methods, where static curricula may not address diverse learning styles, AI tools provide dynamic and multimodal experiences, helping bridge gaps in student understanding. Moreover, generative AI and LLMs foster interactive learning environments, aiding students in acquiring problem-solving skills compared to traditional education environments. Figure 2 illustrates the overlapping aspects between new-age and traditional learning methods.

2.1 Traditional education patterns

Traditional education is often characterized by teacher-led instruction, standardized content delivery, and limited personalization. According to a review by Hattie [14], traditional education approaches have shown varying degrees of effectiveness depending on factors like teaching quality, student engagement, and classroom environment. However, critiques of this system often highlight its rigidity, limited student autonomy, and failure to handle the individual learning styles. Piaget's [15] theory of cognitive development and Vygotsky's [16] social constructivist theory laid much of the groundwork for these traditional models, emphasizing structured learning environments and the importance of social interaction in education. Despite its widespread use, several studies have shown that traditional methods often result in disengaged students and limited long-term retention of knowledge. Research done by Govaerts et al. [17] indicated that passive learning methods and rote memorization can contribute to decreased student motivation, as they often fail to provide the flexibility or real-world application that many students require.

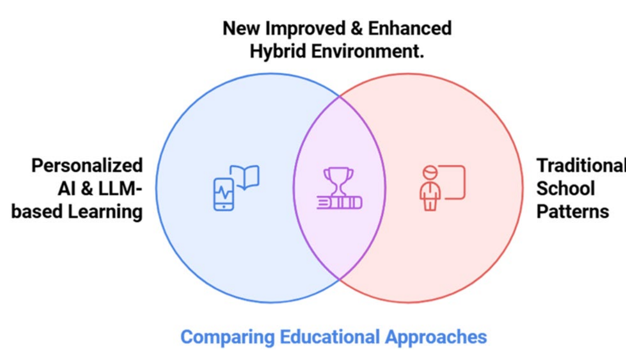


Fig. 2 Overlapping aspects of new-age and traditional learning

2.2 AI-based learning

Recent studies highlight the integration of AI in mathematics education, focusing on adoption, literacy, and teacher knowledge. One study reveals that performance expectancy significantly influences preservice teachers' intention to use AI chatbots, emphasizing the need for targeted workshops to enhance adoption [18]. Latent Profile Analysis of AI Literacy and Trust in Mathematics Teachers identifies five distinct profiles of AI literacy and trust, showing that higher literacy and trust levels correlate with increased AI dependency, which may negatively impact essential skills like critical thinking and problem-solving. The study calls for a balanced approach to AI integration to prevent skill erosion [19]. *Teachers' AI-TPACK: Exploring the Relationship between Knowledge Elements* introduces the AI-TPACK framework, examining the complex interplay between technological, pedagogical, and content knowledge elements. The study finds that technical knowledge elements strongly predict AI-TPACK effectiveness, emphasizing the need for comprehensive teacher training to enhance AI integration sustainably. Collectively, these studies highlight the importance of fostering positive AI attitudes, balancing AI use, and providing targeted professional development to optimize AI's role in mathematics education [20].

On the other hand, AI-based learning systems, particularly those driven by LLMs (like GPT-4 or BERT), offer an alternative that addresses many of the shortcomings of traditional education. Personalized AI-driven learning platforms use algorithms to adapt to individual student needs, adjusting the difficulty of material, pacing, and instructional style in real-time. According to Zawacki-Richter et al. [21], AI in education has been shown to significantly improve learning outcomes by providing tailored feedback and immediate support. A meta-analysis by Wang et al. [22] demonstrated that students using AI-based platforms showed improved knowledge retention and engagement compared to those in traditional learning environments. One key advantage of AI-based learning lies in its adaptive nature. As described by Luckin et al. [23], AI can dynamically adjust the learning path based on a student's performance, ensuring a continuous challenge while minimizing frustration. This personalized approach caters to a broader range of learning styles and paces, making it more inclusive. Additionally, AI models like LLMs have revolutionized learning content generation by understanding natural language queries and providing detailed, contextualized responses that enhance the learning experience. These findings underscore the need for balanced AI use and targeted professional development to enhance teaching outcomes.

2.3 Effectiveness of AI-based learning vs. Traditional methods

Several studies suggest that personalized AI-based learning has the potential to outperform traditional methods. Holstein et al. [24] found that AI-based tutoring systems improved student performance by providing immediate feedback and personalized learning strategies. Another study by Roll and Wylie [25] emphasized that AI-driven adaptive learning systems can improve conceptual understanding, particularly in subjects like math and science, where individualized pacing is critical. However, the transition to AI-powered learning is not without challenges. Critics argue that the reliance on algorithms may result in students losing the social and emotional benefits of classroom learning, a factor well-documented in Bandura's social learning theory [26]. Additionally, concerns about accessibility and equity in AI-driven education are often raised, as not all students have equal access to the technology required to benefit from these systems.

Table 1 highlights the differences between traditional education methods and AI-based learning systems, focusing on key aspects such as personalized learning, adaptability, scalability, and resource efficiency. It demonstrates how AI-based systems offer enhanced customization and interactivity compared to conventional approaches.

2.4 Case studies on the integration of ChatGPT in education

In recent years, the integration of ChatGPT and other AI tools into educational environments has sparked significant interest and debate. Various case studies have explored its impact across different educational contexts, from enhancing student participation in online discussions to reshaping traditional performance evaluation methods. These studies highlight both the potential benefits and challenges of incorporating AI into teaching and learning, offering valuable insights for future educational practices and policy development. The Table 2 summarizes key findings from several such case studies, illustrating the diverse applications of ChatGPT in education.

Table 1 Comparison of traditional education patterns and AI-based learning systems

Aspect	Traditional education	AI-based learning
Characteristics	• Teacher-led instruction • Standardized curricula • Limited personalization	• Personalized learning paths • Dynamic content generation • Real-time feedback and adaptation
Theoretical foundations	• Piaget's Cognitive Development Theory (1970) [15] • Vygotsky's Social Constructivist Theory (1978) [16]	• Luckin et al. [23]: Adaptive learning pathways • Zawacki-Richter et al. [21]: AI-tailored instructional design
Strengths	• Structured learning environments • Emphasis on social interaction (Vygotsky)	• Tailored pacing and material difficulty • Multimodal and interactive learning • Enhanced knowledge retention
Weaknesses	• Rigid and inflexible • Limited accommodation for diverse learning styles • Reliance on rote memorization Govaerts et al. [17]	• Potential loss of social and emotional benefits [26] • Equity and accessibility issues •Dependence on algorithms
Effectiveness	• Hattie [14]: Effectiveness varies with teaching quality and engagement • Govaerts et al. [17]: Passive methods reduce motivation	• Wang et al. [22]: Improved retention and engagement • Holstein et al. [24]: Immediate feedback enhances performance
Motivation & engagement	• Often low due to lack of flexibility and real-world application Govaerts et al. [17]	• High due to personalization and real-time adaptation
Retention of Knowledge	• Limited long-term retention due to static curricula	• Enhanced retention through customized content and pacing Wang et al. [22]
Application to Subjects	• Limited flexibility in complex, dynamic subjects	• Effective for math, science, and problem-solving-oriented subjects Roll and Wylie, [25]
Social Interaction	• High social interaction in classroom settings	• Potentially reduced social interaction [26]

Table 2 Overview of Case Studies on ChatGPT Integration in Educational Settings

Study title	Abstract & conclusion	Key findings	Challenges/concerns
"A Case Study Investigating the Utilization of ChatGPT in Online Discussions" [27]	This study explored the integration of ChatGPT into asynchronous online discussions, analyzing student log data and their perspectives. The findings revealed that ChatGPT significantly enhanced overall participation and encouraged constructive conversations, fostering critical thinking and community-building among students. However, challenges such as superficial responses and disconnection during discussions were also noted. In conclusion, the study highlighted ChatGPT's potential to improve student engagement and knowledge exploration, while also recognizing its limitations. It suggests that future studies should expand the sample size, explore the role of AI literacy, and investigate the effects of ChatGPT on student participation compared to those not using the tool	- Increased participation and engagement - Enhanced critical thinking and knowledge exploration - Positive student views on ChatGPT fostering constructive conversations	- Superficial responses - Disconnection during discussions
"Artificial Intelligence (AI) in Education: A Case Study on ChatGPT's Influence on Student Learning Behaviors" [28]	This research examined ChatGPT's impact on learning behaviors at Ho Chi Minh City University, demonstrating its widespread use by students for academic tasks like idea generation and assignment completion. Despite its advantages in accessibility and efficiency, concerns about overreliance, reduced creativity, and ethical issues like plagiarism emerged. In conclusion, while ChatGPT significantly impacted student learning by making tasks more accessible, its integration requires a balanced approach. Ethical concerns need to be addressed through policies, and future research should focus on understanding the long-term effects of AI tools on learning outcomes and ethical frameworks for their implementation	- Improved learning efficiency - Recognition of ChatGPT's role in academic tasks - Concerns about plagiarism	- Overreliance on AI - Academic dishonesty - Reduced creativity

Table 2 (continued)

Study title	Abstract & conclusion	Key findings	Challenges/concerns
"ChatGPT and the EFL Classroom: Supplement or Substitute in Saudi Arabia's Eastern Region" [29]	This study compared EFL learners' satisfaction with teacher-mediated versus ChatGPT-assisted writing tasks, finding that students were more satisfied with teacher involvement in most aspects, particularly in providing interactive opportunities. ChatGPT was mainly valued for ease of use and grammar corrections. The conclusion emphasizes that while ChatGPT can supplement the learning process, it cannot replace teachers in fostering engagement, motivation, and classroom management. It suggests that teachers can use ChatGPT as a tool for grading and feedback but should maintain a central role in facilitating learning and motivation	- Positive feedback on ChatGPT's role in assignment review - EFL learners preferred teacher involvement for motivation - ChatGPT helps with grammar and vocabulary improvement	- Lack of engagement and motivation - Limited classroom management capabilities
"Time to Revisit Existing Student's Performance Evaluation Approach in Higher Education Sector in a New Era of ChatGPT" [30]	This study explored the implications of ChatGPT on performance evaluation methods in higher education. It tested ChatGPT-generated assignments against plagiarism detection tools like Turnitin, finding that the AI-generated work easily passed these checks. The study concluded that traditional performance-based evaluation methods are no longer sufficient, as students could claim AI-generated work as their own. It recommends that higher education institutions revise their evaluation methods to account for AI tools like ChatGPT and develop new methods to validate authorship and maintain academic integrity	- ChatGPT can generate high-quality academic work - It passes plagiarism checks from tools like Turnitin - Current performance-based evaluation methods are insufficient in the AI age	- Difficulty in detecting AI-generated work - Students claiming AI-generated work as their own - Lack of authorship validation methods

Table 2 (continued)

Study title	Abstract & conclusion	Key findings	Challenges/concerns
"Integrating ChatGPT in Education and Learning: A Case Study on Libyan Universities" [31]	This study assessed the potential for integrating ChatGPT into education at Libyan universities, revealing strong motivation among students and faculty to use the tool, particularly for research and study purposes. Despite some technical challenges, like unstable internet connections and unfamiliarity with ChatGPT, the study concluded that there is widespread support for its adoption. The study suggests that future integration should be accompanied by clear guidelines and training programs to ensure effective use and prevent overreliance, which could hinder creativity and academic integrity. Furthermore, awareness programs and regulations on AI tool usage should be implemented to ensure its responsible integration into education	- Strong motivation to use ChatGPT in research and studies - Widespread consent for integrating ChatGPT into education - Mobile apps are the most common access point	- Technical issues like unstable internet - Unfamiliarity with ChatGPT - Concerns about overreliance and reduced creativity

Fig. 3 Developing effective research questions and methods

Table 3 Research question's

Research questions (RQs)	Description
RQ1	Do LLM-based educational integration improve student engagement and retention compared to traditional methods?
RQ2	Can LLMs contribute to the emotional and social development of students in addition to academic success?
RQ3	Can LLMs be trusted to develop a fair and rigorous examination environment across different academic levels?
RQ4	How can LLMs be utilized for real-time monitoring of student progress while ensuring privacy and data security?
RQ5	What challenges and ethical considerations have arisen from LLM's Models usually implement in the field of AI in Education-today
RQ6	To what extent can LLMs accurately simulate real-world teaching experiences, such as Socratic questioning?

3 Methodology design

This study aims to analyze the influence of AI (LLMs Agents) [32] in the field of education, instruction, administration, learning, evaluating and guiding. This study uses a detrimental technique, using various forms of available data and resources of past investigation and current trends in education. Snyder proposed that a systematic or semi-systematic literature review, or a review of secondary data, gives a more in-depth knowledge of the research phenomena [33]. Because only studies, including meta-analyses, that have been done on the topic support the identification, analysis, understanding, and synthesis of the ways in which AI has affected and impacted education, this approach guarantees that the study is empirically based, or evidence backed. To evaluate the various approaches, a qualitative study design that incorporates topic analysis and qualitative content is often used. To draw findings and draw inferences for descriptive research, thematic and content analysis include carefully analyzing each text and detecting recurrent themes from a study of several texts [34]. Given the goal of this study, which is to evaluate the impact of personalized AI, LLM-based learning & education, the research methodology and approach are adequate.

Figure 3 illustrates the process of formulating effective research questions and selecting appropriate methods, emphasizing the importance of clarity, relevance, and alignment with research objectives. It showcases a step-by-step framework to guide researchers in achieving precision and methodological rigor.

3.1 Research question's

This review is organized to address six critical research questions (RQs) that investigate the influence of personalized AI and LLM-based learning systems on student knowledge enhancement, contrasting these outcomes with those of traditional educational patterns. Table 3 presents a detailed outline of these research questions, focusing on aspects such as adaptability, knowledge retention, engagement, accessibility, and overall learning effectiveness.

3.2 Search strategy

To ensure a comprehensive literature search, the following academic database were selected IEEE Xplore, Scopus, Web of Science and Google Scholar is searched using keywords and search strings to find publications from various journals

Table 4 Inclusion and exclusion criteria parameters

Criteria Type	Detail
Inclusion criteria	• Papers published between 2020 and 2024 • Peer-reviewed research articles, conference papers, book chapters and journal articles • Studies focusing on the application of LLM's based AI technology in education • Papers published in English
Exclusion criteria	• Studies that are outside the educational sector • Non-peer-reviewed materials • Papers focus primarily on fact proven or experientially proven studies that ensure a focus on relevant survey • Skipping duplicate articles across multiple digital platforms and selecting Articles using relevant keywords and information

Table 5 Quality assessment criteria for evaluating selected studies

Assessment criteria	Questions for evaluation
Relevance	Does the paper contribute to answering the research questions?
Methodological rigor	Is the research methodology robust and well-explained, including research design, data collection and analysis techniques?
Contribution to knowledge	Does the Paper offer new insights or innovation within the field?
Citation impact	Has the paper been cited widely within the academic community indicating its relevance?

that have examined the effects of artificial intelligence on education. After using an elimination procedure, a total of over forty articles were chosen through inclusions and exclusion Criteria. The search strategy involved developing structured keyword combinations focusing on intersection of Agentic-LLM in the education field. The search keywords included.

- "Comparative Analysis of Traditional education and Learning Techniques
- "AI in education" OR "Generative AI in Education"
- "Multimodal LLMs in Education" OR "Future of Education in the age of AI"
- "AI learning techniques in Education enhancement"
- "Social and mental health care of students in Education"

The inclusion period was limited to papers between 2020 and 2024, ensuring that only the most recent and relevant survey papers in the field of AI in education will be considered.

Search Query Example

("Artificial Intelligence") OR ("Education") AND ("Systematically Comparison")

3.3 Inclusion and exclusion criteria

To ensure the selection of high-quality and relevant studies, specific inclusion and exclusion criteria were established. These parameters focused on the study's relevance to personalized AI and LLM-based learning, methodological rigor, and alignment with the research objectives. Studies were included based on their emphasis on knowledge enhancement and excluded if they lacked empirical evidence or focus on traditional education systems. Table 4 summarizes the detailed inclusion and exclusion criteria applied in this review.

3.4 Quality assessment

To maintain the integrity and relevance of the review, each selected paper was evaluated using a set of predetermined quality assessment criteria. These criteria were designed to assess the relevance, methodological rigor, contribution to knowledge, and citation impact of each study. This process ensured that only high-quality research aligning with the objectives of this review was included. Table 5. Quality Assessment Criteria for Selected Studies summarizes the key questions used for evaluation across various dimensions.

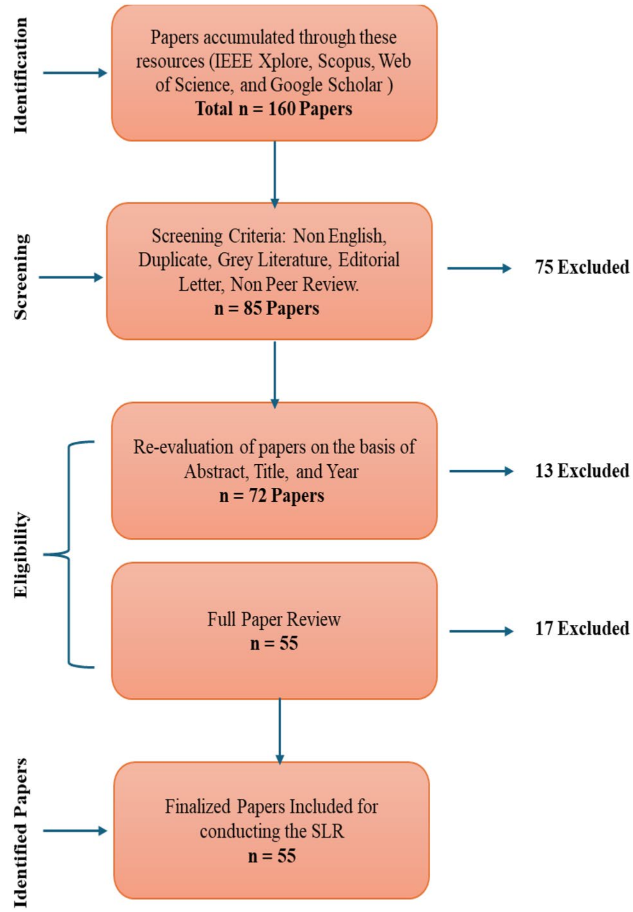


Fig. 4 Study selection process flow

SLR followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines to ensure a transparent and replicable election process. Figure 4 illustrates the selection process for the studies included in this literature review. Starting with an initial set of 160 papers gathered from various academic databases, the figure shows the step-by-step screening and selection criteria that led to the final inclusion of 55 papers. The process involved removing non-English articles, duplicates, grey literature, editorial letters, and non-peer-reviewed sources, followed by a detailed review of abstracts, titles, and publication years.

4 Discussion

Thus, the blending of digital technologies into higher education has become a vital subject of inquiry, offering valuable insights for academic exploration and academic based implementation. Large Language Model (LLM)-based educational integrations have emerged as transformative tools in modern education. Unlike traditional methods, where access to mentor and informative resources is limited by location and schedule, Agentic LLMs environment democratizes education by providing interactive, personalized, and instantaneous feedback to students. According to Yarychev and Mentsiev [35], digital education broadens accessibility, enabling a "knowledge society" where learning is no longer constrained to physical spaces or other limitations.

Thus, students can use LLMs ethically to expand their knowledge in any field by engaging in prompt or natural language conversations like those used by humans. Without constantly asking teachers, they got entangled on some questions. The flowchart below depicts the new Agentic LLMs-based top-view progression of a student's request through an agent-based system designed to generate appropriate responses or actions. Figure 5 presents a flowchart demonstrating the operation of an Agent-based LLM.

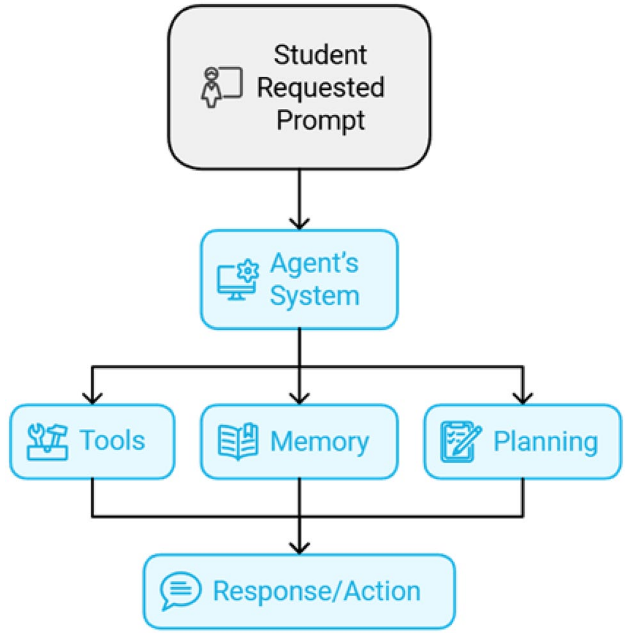


Fig. 5 Flowchart showing how agent based LLM works

- Student Requested Prompt: The Student Requested Prompt represents the student's input, such as a question, query, or task. In an educational setting, this could be a request for information, an explanation of a subject, or an evaluation question. This could range from requesting a summary of a course to seeking assistance with a specific problem or topic.
- Agent's System: Once the student's prompt is received, it is passed to the agent, which handles the decision-making process. This system is central to LLMs in education because it determines how the system will respond to the students' request. The agent's role is to manage the tools, memory, and planning involved in generating a suitable answer or action.
- Tools: The tools block refers to the available resources or external systems that the agent can use to provide an answer or perform a task. In educational settings, these could include access to a database of educational content, computational tools for solving problems (small example like a calculator for math problems), or other systems integrated with the LLM to facilitate interactive learning (e.g., a code compiler or virtual classroom tools). Tools are used to support the agent's ability to generate an informative response.
- Memory: This represents the storage or access to previous interactions, context, and personalized information. In an educational scenario, memory allows the agent to remember past conversations, adapt responses based on previous learning, and personalize the experience for the student. For example, an LLM-based tutor could recall a student's progress in a course, adjust its explanations based on previous misunderstandings, and track progress over time.
- Planning: The planning component focuses on how the agent organizes its responses. Based on the students' request, the system must plan how to generate an answer. In the context of LLMs, this could involve deciding whether to search for information, explain a concept in detail, give an example, or ask the student follow-up questions to clarify their request. This is particularly crucial in more complex queries where the agent needs to break down the task into smaller steps or provide structured explanations.

4.1 Do LLM-based educational integration improve student engagement and retention compared to traditional methods?

AI-powered Large Language Models (LLMs) are transforming education by providing dynamic, personalized learning experiences that overcome the limitations of traditional one-size-fits-all approaches. Unlike traditional strategies, which frequently fail to effectively engage diverse learners, LLMs use advanced natural language processing and adaptive algorithms to tailor content to each student's needs. For example, personalized feedback, curriculum pacing, and real-time assessments improve both engagement and comprehension, with studies indicating a 32% increase in student retention rates in AI-integrated learning environments [8].

LLMs also promote inclusivity by supporting multiple languages and learning styles, which addresses the global challenge of equitable education. According to a World Economic Forum [9] survey conducted in 2022, 64% of students who used AI tools outperformed their peers in traditional settings in terms of learning outcomes. These capabilities position LLMs as transformative educational tools that promote deeper learning, critical thinking, and long-term academic success. Agentic LLM's Improved Impact's on Student Learnings:

4.1.1 Personalized learning

LLM based AI-enabled educational systems or environments powered by Multiple Agent's are sufficient at analyzing student progress in real-time through continues student interactions, allowing both teachers and AI agents to adapt educational content to meet specific needs. For instance, a student struggling with foundational mathematical concepts can receive simpler explanations and additional exercises to solidify their understanding, while an advanced student can access more challenging material to stay engaged. This level of customization goes far beyond the capacity of traditional classroom settings, where teachers often face the challenge of addressing diverse learning paces and styles within a limited timeframe. Which allows teachers and smart LLMs agent's capabilities to tailor educational content according to the learner's specific needs, strengths, and weaknesses. This personalization fosters a deeper engagement with the learner's mind and keep students motivated and reduces disengagement due to irrelevant or overly complex material [36, 37]. Research shows that personalized learning increases student retention rates and improves academic outcomes, making it a vital component of modern education. These systems are not limited to academic content; they can also integrate learning resources for socio-emotional skills, encouraging holistic growth. Overall, by combining real-time analytics and personalized strategies, LLM-driven learning systems pave the way for a more inclusive and impactful education system.

4.1.2 Real-time feedback and adaptation

One of the most transformative aspects of LLM-driven educational systems is their ability to provide real-time feedback and adaptation. Unlike traditional classroom settings where feedback is often delayed due to time constraints or the overwhelming student-to-teacher ratio, these LLMs based systems can instantly analyze responses and provide actionable suggestions. For instance, when a student submits an essay or answers a complex problem, the system not only identifies errors but also explains the reasoning behind corrections, offering immediate guidance. This type of instant feedback has been proven to significantly enhance learning outcomes by addressing misconceptions at the moment they arise. Due to data access of student's to LLMs can instantly analyze student's responses and provide immediate, actionable feedbacks, enhancing student overall involvement in the learning process. This feedback mechanism is very difficult in the traditionally driven education environment where in a classroom of 1 or 2 teachers are there to correct you but due to LLM based Agentic flow each users can now get instant feedback and direction to resolve their issues and concerns about any topic's in any subject, which lead to a good retention of engagement in study [36, 37]. Moreover, real-time feedback isn't limited to academic tasks. These systems also offer emotional and motivational support, encouraging students when they feel stuck or overwhelmed. This combination of cognitive and emotional guidance transforms the learning experience into a more engaging and supportive process. The integration of these systems into education bridges the gap between traditional and modern learning, creating a more dynamic and responsive environment where students can thrive.

4.1.3 Increased motivation through gamification

Not everyone is a fan of old-school textbook based learning, where in some research it says people learn and understand more through visuals so, many AI-powered educational tools include visual gamified elements by introducing challenges, levels, and rewards, these tools transform mundane learning tasks into exciting and engaging activities. For example, a history lesson might become an interactive game where students "unlock" historical events by answering questions correctly, while a math exercise could involve earning points while solving a mystery based math problem where agents provide clues to students time to time so that student can solve it [36]. Visual and interactive elements are especially effective for students who struggle with text-heavy content. Research shows that 65% of individuals are visual learners, meaning they absorb information more effectively when it's presented through images, videos, or other visual formats. LLM-based systems capitalize on this by integrating vibrant graphics, animations, and even immersive simulations to teach complex concepts. These engaging methods make learning feel less like a chore and more like an adventure, fostering curiosity and enthusiasm among students. Additionally, gamification provides immediate gratification through rewards, which keeps students motivated to progress. This motivational boost isn't limited to academic content; it also promotes soft skills like teamwork, problem-solving, and time management. By merging gamification with personalized learning pathways, LLM-based systems create a holistic educational experience that caters to both academic and emotional growth.

4.1.4 Continuous learning pathways

Education should not end when a student leaves the classroom. One of the most remarkable contributions of AI-powered systems, particularly those leveraging LLMs, is their ability to facilitate continuous learning pathways. These systems offer a seamless transition between different stages of education, ensuring that students are always equipped with relevant and updated knowledge. For instance, an AI system can track a student's progress throughout their academic career, recommending advanced courses, supplementary resources, or even career-oriented skill development programs as they advance. This flexibility greatly enhances retention, as students are not confined to static curricula [36, 37]. Moreover, these systems can identify gaps in a student's understanding and address them proactively, preventing the accumulation of knowledge deficits. For example, a student preparing for a standardized test can use the AI system to review past topics, identify weak areas, and receive customized study plans to improve. Beyond academic growth, continuous learning pathways foster a mindset of lifelong learning. In today's rapidly changing world, where industries evolve at an unprecedented pace, this adaptability is crucial. LLM-powered systems prepare students not just for exams but for real-world challenges, equipping them with the skills and knowledge needed to thrive in a dynamic global landscape. By creating an environment where learning never stops, these tools ensure that students remain engaged, curious, and prepared for the future.

4.2 Can LLMs contribute to the emotional and social development of students in addition to academic success?

The integration of Large Language Models (LLMs) in education presents a transformative opportunity to enhance not only academic success but also emotional, social, and cultural development. While traditional educational systems tend to prioritize academic performance, often overlooking students' emotional and social needs, LLMs have the potential to address these gaps by offering personalized support, promoting communication skills, and fostering mental well-being. Through natural language processing (NLP) and adaptive interactions, LLMs can facilitate meaningful dialogue with students, providing them with a platform to express their thoughts and emotions. For example, AI-driven emotional intelligence tools can assess sentiment and generate tailored responses. A 2023 study by Chen et al. [38] found a 40% increase in students' self-reported emotional well-being when using AI-powered conversational systems in educational contexts. However, while these findings are promising, it is important to note that further research is necessary to fully understand the underlying mechanisms that contribute to these improvements and to validate such claims.

In addition to emotional support, LLMs can foster the development of social skills by encouraging collaborative learning through virtual group discussions and role-playing scenarios. Studies have suggested that platforms incorporating LLMs have improved peer communication skills by up to 28% [39]. However, these claims, though notable, require more rigorous and longitudinal studies to confirm their robustness and long-term impact. By simulating real-world scenarios and promoting empathy, LLMs have the potential to encourage emotional resilience, cultural awareness, and interpersonal growth. These contributions suggest that LLMs may play a key role in shaping not only academic success but also the emotional and social dimensions of education. Nevertheless, the impact of LLMs on these aspects of development must be further substantiated through empirical research, as their exact effects and the mechanisms involved remain unclear.

4.2.1 Emotional support and well-being

LLMs offer significant promise in supporting students' emotional well-being by understanding and responding to emotional cues. With increasing pressures from academic demands, social expectations, and personal challenges, students often face emotional difficulties such as stress, anxiety, and isolation. The integration of LLMs into educational systems could offer a solution by providing consistent, empathetic, and accessible emotional support. These AI systems, using NLP, can simulate empathetic conversations and provide a space for students to express their emotions. While there is potential for LLMs to serve as a first line of defense for students hesitant to seek help due to stigma, further research is needed to understand the specific ways in which LLMs can best support students and whether they can provide the same depth of support as human counselors.

- Personalized Emotional Support: LLMs can engage in non-judgmental conversations, allowing students to express their emotions freely without the fear of stigmatization. This is particularly beneficial for students in need of mental health support but who are hesitant to approach human counselors [40, 41]. AI-powered systems, such as LLM-based web platforms, can identify distress signals in students' language and provide tailored, calming responses. These interactions are powered by vast datasets that allow LLMs to recognize and address various emotional cues, such as stress or frustration. However, while these interactions may be beneficial, the degree to which LLMs can effectively address complex emotional issues and provide sufficient support remains a topic for further investigation.
- Real-Time Feedback: Studies have indicated that AI-driven systems can provide real-time feedback during emotionally charged situations, such as before exams or during group projects, helping students navigate conflicts or uncertainties [40]. This could contribute to promoting emotional resilience, helping students stay focused on their goals while managing stress. Nevertheless, while initial findings suggest that LLMs can provide real-time interventions, more evidence is needed to determine the effectiveness and reliability of these interventions, especially in the long term [42].

LLMs may play a complementary role to traditional counseling services by offering accessible emotional support. However, further research is necessary to fully evaluate the effectiveness and sustainability of such interventions and their role in preventing emotional burnout and promoting students' overall well-being.

4.2.2 Social development through AI-facilitated interactions

The evolving landscape of education emphasizes collaboration and the development of social skills. Traditional education systems often emphasize individual achievement and competition, while LLMs offer opportunities for collaborative learning through virtual discussions, group activities, and problem-solving tasks. Research has suggested that platforms incorporating LLMs can enhance peer communication by up to 28% [39]. However, more in-depth studies are needed to verify these claims and explore the specific ways LLMs contribute to social development.

LLMs are also used in gamified learning environments that encourage teamwork, critical thinking, and shared responsibility. These platforms, including educational games like Duolingo or collaborative coding tools, foster team-based learning, promoting the development of skills like negotiation, compromise, and leadership. However, while these gamified tasks have been shown to enhance collaboration, the extent to which they contribute to long-term social skill development requires further empirical analysis. Additionally, further evidence is needed to evaluate whether these skills transfer effectively to real-world contexts, such as workplace environments.

4.2.3 Cross-cultural sensitivity and communication

In a globalized world, the ability to communicate effectively across cultures is a vital skill. LLMs, trained on diverse datasets, can help foster cross-cultural communication by providing multilingual support and culturally sensitive recommendations. In language learning platforms, LLMs not only assist with grammar and vocabulary but also expose students to cultural nuances, idiomatic expressions, and region-specific communication practices [42, 43]. This helps broaden students' perspectives and improve their understanding of diverse worldviews.

LLMs also help break down language barriers by offering AI-powered translation tools that enable students to communicate and collaborate with peers from different linguistic backgrounds. By fostering empathy and understanding of cultural subtleties, LLMs support students in developing inclusive communication strategies [44]. However, while these tools can enhance cross-cultural communication, the depth of cultural understanding and the potential limitations of AI in addressing complex intercultural nuances need to be carefully evaluated.

Moreover, LLMs can simulate real-world scenarios that require students to interact with individuals from different cultural backgrounds, providing a safe space for practicing cross-cultural communication skills. While this is a promising avenue for global learning, future advancements in LLM technology will need to ensure that they are able to provide context-sensitive, culturally relevant recommendations to better prepare students for success in multicultural environments. The evolving capabilities of LLMs in this area warrant further exploration to fully understand their impact on students' global communication skills.

4.2.4 Personalized learning and emotional support

LLMs also offer personalized learning and emotional support in ways that can adapt to both academic and emotional needs. By analyzing student performance, preferences, and emotional cues, LLMs can create customized learning paths tailored to individual needs. For example, if a student struggles with a particular concept, an LLM-powered system might offer supplementary materials, different explanations, or even shift the learning modality (from text-based to video-based) based on the student's preferences [41]. Additionally, emotional check-ins using multimodal data (such as audio or video analysis) allow LLMs to detect shifts in a student's emotions. If a student shows signs of frustration or disengagement during an online session, the AI can intervene with supportive messages, suggest a break, or propose a simpler activity to improve emotional well-being [41]. Personalized feedback, which emphasizes effort and perseverance, encourages a growth mindset, further supporting emotional and academic development. However, while these features offer promising support, the overall impact of LLMs on both emotional development and academic success requires further empirical validation.

4.3 Can LLMs be trusted to develop a fair and rigorous examination environment across different academic levels?

As societies are evolving with the integration of artificial intelligence (AI), the promise of Large Language Models (LLMs) like ChatGPT, Bard etc. to revolutionize examination environments has sparked a mix of optimism and caution. The potential for LLMs to automate grading, generate exam content, and personalize assessments offers a significant leap forward, but it also raises critical questions about fairness and rigor in academic evaluation. Can these systems be trusted to create an equitable examination environment across diverse academic levels, free from biases and inconsistencies?

The pursuit of fairness in AI-driven examinations is far from straightforward. While research by Mitra [45] highlights the promise of multi-agent systems in generating balanced and rigorous exam content, findings from Mariyono [40] and Rane [42] emphasize the persistent risks of algorithmic bias and hidden inequalities. Furthermore, studies such as Benitta et al. [43] suggest that hybrid systems integrating human oversight and AI may offer a pathway to equitable grading, while Liladhar et al. [46] confirm AI's ability to apply standardized test protocols consistently. These insights reveal a dual narrative: one of transformative potential and another of unresolved challenges in ensuring transparency and fairness. Below papers delves into a comparative analysis of recent studies to explore whether LLMs can be trusted to develop a fair and rigorous examination environment. Table 6 provides an analysis of studies examining the ability of LLMs to establish fair and rigorous examination environments across various academic levels, highlighting key findings and unique.

For the future, researchers agree on the importance of proactive measures to build trust in AI-driven examination environments. These include implementing rigorous bias audits, increasing transparency in algorithmic decision-making, and adopting adaptive systems that can accommodate diverse student populations. The development of ethical guidelines and collaboration between AI developers, educators, and policymakers will be crucial to ensuring that LLMs contribute positively to fair and rigorous academic evaluations.

4.4 How can LLMs be utilized for real-time monitoring of student progress while ensuring privacy and data security?

Applying Large Language Models (LLMs) in the classroom offers a revolutionary way to track the growth of learners in real time. When it comes to performance analysis, providing tailored feedback, and protecting data privacy, these models are unmatched. But implementing them requires striking a careful balance between utilizing technology's promise and taking important privacy and ethical issues into account. In contrast to traditional methods, which typically take a generic strategy, AI-powered Large Language Models (LLMs) offer dynamic, personalized learning experiences by customizing content to each student's needs, making their integration into educational systems transformative, especially in terms of improving student engagement and retention.

Table 6 Analysis of LLMs' capabilities in creating fair and rigorous examination environments across academic levels

S. No	Author	Study focus	Fairness observations	Rigor insights	Key contributions
1	Mariyono et al. [40]	Ethical concerns in AI-enhanced learning	Algorithmic biases risk unfair grading. Recommendations include bias audits.	Highlighted potential for real-time adaptive assessments tailored to students.	Strong emphasis on inclusivity and equity, calling for continuous dataset diversification and teacher oversight to ensure fairness.
2	Rane et al. [42]	Education 4.0 and 5.0: AI-driven learning systems	Discussed how AI can perpetuate hidden biases, despite its scalability.	Explored intelligent tutoring systems for rigorous, adaptive learning.	Proposed federated learning to secure privacy while ensuring diverse datasets are used to reduce bias and achieve rigorous assessments.
3	Mitra et al. [45]	Generative AI in teaching and assessment	Multi-agent systems showed promise in generating balanced exam content.	Significant improvements in accuracy and reduced hallucination rates.	Introduced Agent Instruct to refine synthetic datasets for rigorous exams. Emphasized creating diverse, unbiased exam questions.
4	Dwivedi et al. [47]	Multidisciplinary perspectives on generative AI	Highlighted transparency concerns in LLMs for examination settings.	Questioned consistency across academic levels due to dataset limitations.	Proposed human-AI collaboration in exam design to balance rigor and fairness. Encouraged feedback loops to refine AI-driven assessments continuously.
5	Benitta et al. [43]	Hybrid educational systems integrating AI for assessments	Showed potential for AI to provide equitable grading across demographics.	Highlighted AI's capability to manage diverse academic standards.	Advocated for AI-human synergy in grading to ensure fairness while maintaining rigorous evaluation protocols across institutions.
6	Pandiaraj et al. [48]	Sentiment analysis and adaptive AI in learning	Found emotional sentiment analysis could add fairness by understanding student needs.	Stressed adaptability of AI in tailoring exam difficulty to individual levels.	Emphasized using sentiment analysis to detect stress during exams, ensuring a fairer and more supportive environment.
7	Ghimire et al. [44]	Technology Acceptance Model (TAM) for Generative AI in the classroom	Discussed educator concerns over opaque grading algorithms.	Highlighted difficulty in aligning AI models with institutional standards.	Proposed localized fine-tuning of AI to match specific academic levels and reduce discrepancies in rigor and fairness.
8	Borović et al. [49]	Comparative analysis of generative AI tools in education	Found ChatGPT and Bard to generate exam content consistently but flagged biases.	Bing Chat proved more rigorous for logical reasoning assessments.	Suggested multi-tool integration for cross-validation of AI-generated exam content.
9	Liladhar et al. [46]	AI applications in examination automation and adaptive learning systems	Confirmed consistent application of standardized test protocols by AI.	Explored dynamic difficulty adjustments using AI.	Emphasized real-time feedback in exams to maintain both rigor and fairness simultaneously. Suggested use of reinforcement learning for continuous improvements.

4.4.1 Personalized learning and real-time monitoring

LLMs can process large volumes of student data, including assignments, quizzes, and interaction logs, to assess progress dynamically. For instance, LLMs can provide tailored feedback to students, enabling real-time identification of learning gaps. This adaptability enhances student engagement and promotes individualized instruction [50].

Moreover, integrating conversational agents powered by LLMs can facilitate continuous student interaction. These agents can track behavioral patterns, provide instant support, and even offer emotional encouragement, which is critical for holistic progress tracking [51].

4.4.2 Data privacy

Ensuring privacy in educational contexts is paramount. Federated learning, an approach where LLMs train on local devices without transmitting raw data, can be instrumental. This method allows institutions to maintain data integrity while benefiting from the LLM's learning capabilities [52]. Additionally, implementing differential privacy mechanisms can protect sensitive student information by adding noise to datasets, ensuring that individual data points cannot be extracted [53].

4.4.3 Security measures

To secure data during LLM operation, encryption techniques such as homomorphic encryption can be employed. This allows data to be processed in an encrypted state, eliminating vulnerabilities during computation. Moreover, adhering to global standards like GDPR ensures compliance with robust security frameworks, safeguarding data against breaches [54].

4.4.4 Ethical considerations

Using LLMs responsibly involves aligning with ethical AI practices, such as transparency and bias mitigation. It is critical to ensure that data usage is well-communicated to stakeholders, including students and parents, fostering trust and minimizing misuse [55].

4.5 What challenges and ethical considerations have arisen from LLM's Models usually taken to implement in the field of AI in Education-today

The integration of Large Language Models (LLMs) like GPT into the field of education has ushered in transformative opportunities. These AI systems possess remarkable capabilities to tailor personalized learning experiences, automate human-like grading processes, and provide round-the-clock tutoring effectively, making education more accessible and scalable than ever before. With their ability to process and generate human-like text, LLMs empower educators and learners alike, fostering adaptive learning paths and addressing individual needs at unprecedented scales.

However, the rapid adoption of LLMs has been leftover with significant challenges and ethical dilemmas. Questions around bias in AI-generated content, data privacy, accountability for erroneous outputs, and the widening of the digital critical hurdles. These issues can be seen in educational contexts, where vulnerable groups, such as children and students from underrepresented communities, are most affected. Addressing these considerations is not only vital for the sustainable implementation of LLMs but also crucial for ensuring equitable and responsible use of technology in learning environments. Table 7 presents the challenges in implementing Large Language Models (LLMs) in education, highlighting key issues such as data privacy concerns, the need for effective integration into existing curricula, and the potential for bias in model outputs. These challenges underscore the importance of thoughtful, ethical implementation of LLMs to enhance educational outcomes while mitigating risks.

4.5.1 Ethical considerations

1. Equity in education

   One of the most critical ethical concerns surrounding the use of LLMs in education is ensuring equity in learning opportunities. AI-powered tools should not disproportionately benefit students who already have greater access to technology, while disadvantaging those from underserved or marginalized communities. Addressing the digital divide is crucial for ensuring that all learners, regardless of their socioeconomic background, have equal access to

Table 7 Challenges in implementing LLMs in education

Challenge	Description	Recommendations
algorithmic bias	LLMs trained on biased or unrepresentative datasets may reinforce harmful stereotypes and inequities in educational content. For example, models may perpetuate gender or racial biases, affecting grading systems or personalized learning. Studies have shown that biased training data can lead to outputs that disadvantage certain demographics [40, 42]	• Implement robust bias detection and correction mechanisms • Ensure transparency in dataset development and AI model architecture • Incorporate diverse perspectives in the development process, including marginalized groups, to create more equitable systems [42]
Data privacy concerns	The use of student data in AI models raises significant privacy issues, particularly in light of regulations like GDPR. Concerns regarding unauthorized access or misuse of sensitive data are prevalent, given the importance of safeguarding student privacy [40, 42]	• Utilize anonymization techniques and ensure data encryption to protect student privacy • Conduct regular audits and ensure AI systems comply with data protection regulations • Implement clear consent protocols to allow parents and students to understand how their data is used [42]
inclusivity issues	Limited access to technology in underserved communities exacerbates educational disparities, particularly for students in rural or low-income areas. This lack of access to modern tools, internet connectivity, or devices means these students cannot fully benefit from AI-driven learning solutions [40]	• Develop affordable, accessible AI tools for diverse socioeconomic backgrounds • Collaborate with policymakers and NGOs to subsidize technology for underserved students • Design AI-powered platforms with low-bandwidth capabilities and offline functionalities to accommodate limited infrastructure [40]
Dependence on technology	Over-reliance on AI tools can weaken teacher-student relationships, which are essential for holistic emotional and social development. Excessive dependence on technology risks diminishing opportunities for empathy and interpersonal skills, which are crucial for students' overall development [40, 42]	• Position AI as a complementary tool rather than a replacement for human interaction • Train educators to integrate AI tools in ways that support, rather than replace, meaningful teacher-student interactions • Focus on creating educational environments that balance digital tools with human-centered pedagogy [42]
Quality and diversity in content	LLMs might generate repetitive or low-quality responses, which can stifle creativity and critical thinking. Furthermore, these models may not reflect the diversity of perspectives necessary for fostering a well-rounded education. Some models may fail to produce nuanced content or overlook key cultural, social, and historical contexts [45]	• Implement content moderation and review systems to ensure the quality of AI-generated material • Use human oversight to ensure that AI content aligns with educational goals and promotes critical thinking • Provide educators with clear guidelines on how to effectively use AI-generated content in their teaching [39]

AI-enhanced educational resources [42]. For instance, adaptive language-learning platforms that require high-speed internet and modern devices may leave rural and low-income students at a disadvantage, leading to academic inequalities over time [40].

- Real-World Implications**:** If these challenges are not addressed, there is a risk that educational inequalities will worsen, further entrenching disparities in learning outcomes across different demographic groups [53].

2. Accountability and transparency

The implementation of LLMs in education requires robust transparency and accountability mechanisms. Decisions made by AI models, whether for grading, content generation, or recommendations, should be clear and traceable. Lack of transparency can erode trust among students, parents, and educators, particularly when models produce biased or incorrect outputs [40, 47]. For example, an automated grading system flagged essays written by bilingual students as plagiarized due to linguistic differences, highlighting the need for careful design and review. Dwivedi et al. [47] emphasizes the need for transparent model architecture and thorough documentation to build trust. Similarly, Khosrawi-Rad et al. [51] underscore the value of systematic reviews and audit trails for conversational agents in education.

- Real-World Implications**:** A lack of accountability can result in a loss of trust in AI systems, potentially hindering the widespread adoption of AI tools in educational settings [51].

3. Effect on originality and creativity

While generative AI tools can enhance learning, there is a risk that over-reliance on such technologies might stifle students' ability to think independently and solve problems creatively [56]. Classroom experiments have shown that students who frequently rely on AI-generated content struggle with tasks requiring original thinking, compared to those who develop ideas independently [6, 45, 47].

- Real-World Implications: Without proper checks, generative AI could lead to a generation of learners with reduced creative skill sets, undermining the broader purpose of education in fostering innovation [47].

4. Adherence to regulations

The deployment of Large Language Models (LLMs) in educational settings must align with existing ethical guidelines, privacy regulations, and pedagogical principles to ensure their responsible use. Compliance with regulations such as the General Data Protection Regulation (GDPR) is critical, as seen in the case of European schools that had to halt AI-based tutoring projects due to inadequate data protection protocols. Adhering to legal and ethical frameworks minimizes the risk of misuse or unintended consequences, ensuring that AI innovations support educational goals. According to Hooda et al. [50], AI tools used for assessment and feedback must prioritize privacy controls, while Zhang and Aslan [52] emphasize the importance of tailoring AI systems to meet regional and institutional legal requirements. To mitigate risks, schools should conduct regular Institutional Review Board (IRB) reviews, implement robust technical controls like data encryption, and establish clear accountability measures.

- Real-World Implications**:** Non-compliance can lead to legal challenges and loss of public trust, potentially stalling AI adoption in education [54].

4.6 To what extent can LLMs accurately simulate real-world teaching experiences, such as Socratic questioning?

AI-powered education systems, particularly those employing Large Language Models (LLMs), have sparked discussions around their computational abilities to simulate complex teaching strategies like Socratic questioning and inquiry-based learning. These approaches, known for fostering critical thinking and student engagement, challenge AI systems to replicate nuanced and adaptive human teaching methods effectively.

1. Enhanced SOCRATIC QUESTIONING with AI

AI systems are well-suited to mimic the structured, iterative nature of Socratic questioning. For instance, LLMs can create follow-up questions that deepen students' understanding and encourage critical analysis. Tools like ChatGPT have demonstrated the capacity to engage learners in multi-turn dialogues by generating adaptive questions based on student inputs [55]. However, LLMs lack the ability to interpret non-verbal cues such as hesitation, frustration, or excitement—key signals in live Socratic interactions, which guide a teacher's adaptive questioning [2].

2. Facilitation of inquiry-based learning

AI excels at scaffolding exploratory learning, providing students with real-time guidance and resources during problem-solving activities. Studies show that AI tools, when integrated into learning environments, enhance engagement by offering data-driven, adaptive support [57]. However, these tools often fail to dynamically adjust to unpredictable student inquiries, limiting their flexibility compared to human educators [2].

3. Hybrid models: augmenting human educators

   Research emphasizes that AI works best in hybrid models, where it supports educators by automating repetitive tasks and analyzing data, allowing teachers to focus on higher-level cognitive and emotional engagement [3]. This collaboration enhances both teaching quality and student learning outcomes, particularly when applied to personalized and adaptive learning frameworks [55].

4. Challenges and ethical considerations

   Despite their capabilities, AI-driven workflows face challenges in replicating the cultural sensitivity, empathy, and ethical reasoning inherent in human teaching. LLMs' reliance on pre-trained datasets can perpetuate biases and provide contextually inappropriate outputs [50]. Additionally, their inability to intuitively navigate complex emotional and cultural dynamics limits their applicability in diverse educational settings [2].

5. Efficacy in feedback and assessment

   AI-powered tools provide immediate, precise assessments and actionable feedback, making them effective in inquiry-based learning contexts. For example, machine learning algorithms have been used successfully to analyze and improve students' performance and self-regulation [58]. However, these systems often lack the depth of human insight required for nuanced pedagogical adjustments [57].

5 Challenges and limitations

Integrating large language models (LLMs) as educational agents has significant potential, but also presents distinct challenges and limitations let's discuss them:

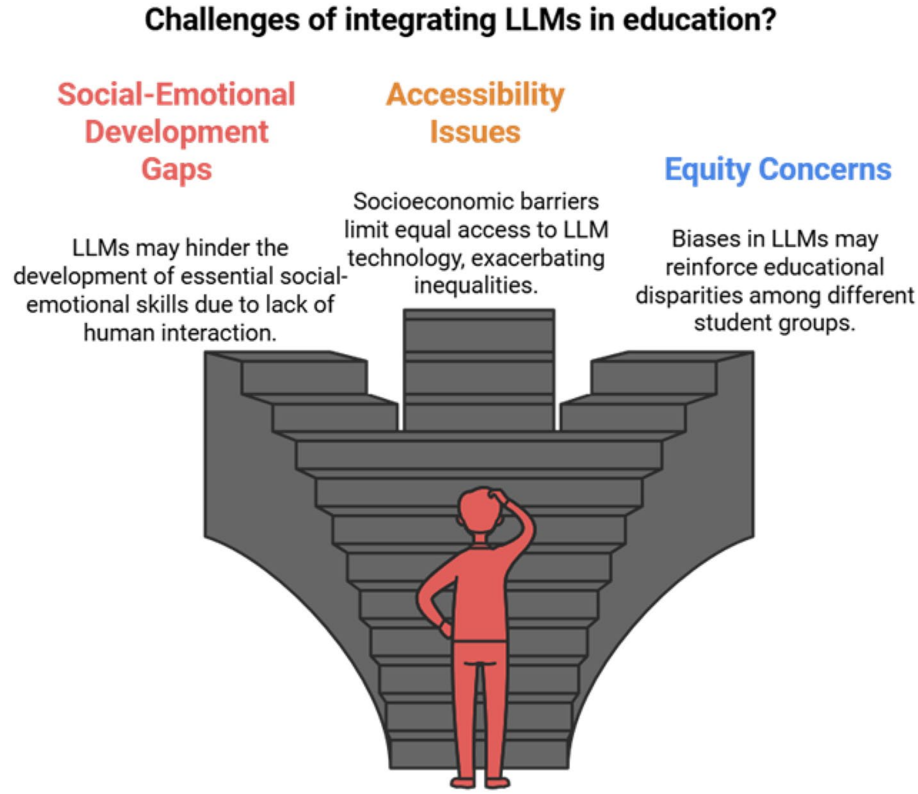
- Social-Emotional Development Gaps: Relying heavily on LLM-based education may diminish opportunities for students to develop essential social-emotional skills due to a lack of authentic human interaction, which is integral for empathy and emotional growth [59]. LLMs cannot fully replicate these nuanced human interactions that foster social bonds and emotional intelligence [10].
- Accessibility Issues: Because not all students have equal access to the equipment, internet connections, or resources required to take advantage of these technologies, socioeconomic hurdles limit the widespread use of LLMs [60]. Furthermore, depending too much on AI-powered resources might result in educational inequalities, particularly in places with limited resources where digital learning infrastructure is still insufficient [61].
- Equity Concerns: Learning tests and instructional materials may contain biases introduced by LLM agents, which might exacerbate already-existing educational disparities. Because LLMs trained on biased datasets may reinforce or magnify these biases, hence disadvantageous student groups, equity concerns arise [62]. Furthermore, if LLM-based tools are not carefully integrated, they may advantage students who are more accustomed to and comfortable with technology, which might lead to a worsening of success disparities [63].

Figure 6 illustrates the current challenges faced in education, including issues like unequal access, outdated curricula, lack of personalization, and the integration of advanced technologies such as AI and LLMs. It highlights barriers impacting both traditional and modern learning systems.

6 Conclusion

The integration of Large Language Models (LLMs) into educational systems presents a transformative opportunity to revolutionize the learning experience. This systematic review of 55 studies published between 2020 and 2024 highlights the significant potential of LLMs in enhancing student engagement, promoting emotional and social development, and improving real-time monitoring of academic progress. Compared to traditional educational approaches, LLMs offer personalized learning pathways that cater to individual needs, allowing for greater student engagement and retention. While LLMs show promise in supporting academic success, their ability to contribute to emotional and social growth presents a compelling case for their inclusion in holistic educational practices. The use of AI-driven emotional intelligence tools can provide students with personalized emotional support, fostering resilience, and

Fig. 6 Current challenges in education



well-being. Furthermore, LLMs facilitate collaborative learning and cross-cultural sensitivity, which are essential skills in an increasingly globalized world. However, despite their advantages, the implementation of LLMs in education raises several critical challenges. Ethical considerations, privacy concerns, and the need for rigorous evaluation to ensure fair and rigorous examination environments remain significant hurdles. While LLMs can help create a fairer, more accessible educational system, further research is needed to ensure their responsible deployment and integration into existing educational frameworks. Moreover, the simulation of real-world teaching experiences by LLMs, such as Socratic questioning, holds significant promise in enhancing interactive and dynamic learning environments. Yet, the extent to which LLMs can replicate nuanced human interactions and pedagogical approaches remains an area for further investigation.

In conclusion, while LLMs have the potential to reshape the educational landscape by offering personalized, inclusive, and accessible learning experiences, their integration must be approached thoughtfully and ethically. Future research should continue to explore the effectiveness of LLMs across diverse academic levels, ensuring that these technologies are deployed in a way that is equitable, effective, and respects the privacy and well-being of students. By addressing the challenges outlined in this review, LLMs can play a pivotal role in advancing education and fostering a more personalized and dynamic learning environment for all students.

Author contributions All authors Sahil Sharma, Puneet Mittal, Mukesh Kumar, Vivek Bhardwaj contributed equally to the manuscript.

Funding Open access funding provided by Manipal University Jaipur.

Data availability No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Wu T, Terry M, Cai CJ. AI Chains: transparent and controllable human-AI interaction by chaining large language model prompts. In *CHI conference on human factors in computing systems*, 2022, pp. 1–22.
2. Nazaretsky T. Teachers. Trust in AI-powered educational technology and a professional development program to improve it. Br J Educ Technol. 2022. https://doi.org/10.1111/bjet.13232.
3. Dai Y. Reconceptualizing ChatGPT and Generative AI as a Student-Driven Innovation in Higher Education. *Procedia CIRP*, 2023.
4. Dimitriadou E, Lanitis A. A critical evaluation, challenges, and future perspectives of using artificial intelligence and emerging technologies in smart classrooms. Smart Learn Environ. 2023;10(12):1–24.
5. Rachha A, Seyam M. Explainable AI in education: current trends, challenges, and opportunities. IEEE SoutheastCon. 2023;2023:232–9.
6. Bernabei M, Colabianchi S, Falegnami A, Costantino F. Students use of large language models in engineering education: a case study on technology acceptance, perceptions, efficacy, and detection chances. Comput Educ Artif Intell. 2023;5:100172.
7. Athanassopoulos S, Manoli P, Gouvi M. The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural classroom. Adv Mobile Learn Educ Res. 2023;3(2):818–24.
8. Smith J, Johnson R, Lee T. AI in education: enhancing engagement and retention. J Educ Technol. 2023;45(3):215–30.
9. Forum WE. The impact of AI tools on global education systems. 2022.
10. Divito CB, Katchikian BM, Gruenwald JE, Burgoon JM. The tools of the future are the challenges of today: the use of ChatGPT in problem-based learning medical education. Med Teach. 2023. https://doi.org/10.1080/0142159X.2023.2290997.
11. VanLehn K. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educ Psychol. 2011;46(4):197–221.
12. Research and Markets. Global Artificial Intelligence (AI) in Education Market (2020 to 2027)- by Technology, Application, Component, Deployment & End-user. 2020.
13. Insights GM. Artificial intelligence (AI) in education market size by technology. 2021.
14. Hattie J. Visible learning: a synthesis of over 800 meta-analyses relating to achievement. Milton Park: Routledge; 2009.
15. Piaget J. Science of education and the psychology of the child. London: Orion Press; 1970.
16. Vygotsky LS. Mind in society: the development of higher psychological processes. Cambridge: Harvard University Press; 1978.
17. Govaerts N, Kyndt E, Dochy F, Baert H. Influence of learning approach and context on workplace learning. J Work Learn. 2011;23(4):225–40.
18. Wijaya TT, Su M, Cao Y, Weinhandl R, Houghton T. Examining Chinese preservice mathematics teachers' adoption of AI chatbots for learning: unpacking perspectives through the UTAUT2 model. Educ Inf Technol. 2024. https://doi.org/10.1007/s10639-024-12837-2.
19. Wijaya TT, Yu Q, Cao Y, He Y, Leung FKS. Latent profile analysis of AI literacy and trust in mathematics teachers and their relations with AI dependency and 21st-century skills. Behav Sci. 2024;14(11):1008.
20. Ning Y, Zhang C, Xu B, Zhou Y, Wijaya TT. Teachers' AI-TPACK: exploring the relationship between knowledge elements. Sustainability. 2024;16(3):978.
21. Zawacki-Richter O, Marn VI, Bond M. Systematic review of research on artificial intelligence applications in higher education where are the educators. Int J Educ Technol High Educ. 2019;16:39. https://doi.org/10.1186/s41239-019-0171-0.
22. Wang X, et al. The efficacy of artificial intelligence-enabled adaptive learning systems from 2010 to 2022 on learner outcomes: a meta-analysis. J Educ Comput Res. 2024;62(6):1568–603. https://doi.org/10.1177/07356331241240459.
23. Luckin R, Holmes W, Griffiths M, Corcier LB. *Intelligence unleashed: An argument for AI in education*. UCL Discovery, 2016.
24. Holstein K, McLaren M, Aleven V. Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 2020, pp. 544–553.
25. Roll I, Wylie R. Evolution and revolution in artificial intelligence in education. Int J Artif Intell Educ. 2016;26(2):582–99.
26. Grusec JE. Social learning theory and developmental psychology: the legacies of Robert Sears and Albert Bandura. Dev Psychol. 1992;28(5):776–86. https://doi.org/10.1037/0012-1649.28.5.776
27. Lin X, Luterbach K, Gregory KH, Sconyers SE. A case study investigating the utilization of ChatGPT in online discussions. Online Learn. 2024;28(2):n2.
28. Nguyen TNT, Lai NV, Nguyen QT. Artificial intelligence (AI) in education: a case study on ChatGPTs influence on student learning behaviors. Educ Process Int J. 2024;13(2):105–21. https://doi.org/10.22521/edupij.2024.132.7.
29. Ahmed MA. ChatGPT and the EFL classroom: supplement or substitute in Saudi Arabia's eastern region. Inf Sci Lett. 2023;12(7):2727–34.
30. Chaudhry IS, Sarwary SAM, El Refae GA, Chabchoub H. Time to revisit existing student's performance evaluation approach in higher education sector in a new era of ChatGPT—a case study. Cogent Educ. 2023;10(1):2210461.
31. Abolkasim E, Hasan M. Integrating ChatGPT in education and learning: a case study on Libyan universities. J Pure Appl Sci. 2024;23(2):19–24.
32. NVIDIA. Introduction to LLM agents. 2023.

33. Snyder H. Literature review as a research methodology: an overview and guidelines. J Bus Res. 2019;104:333–9.
34. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. In: *Lecture Notes in Computer Science*, 2019; 481–493.
35. Yarychev NU, Mentsiev AU. Impact of digital education on traditional education. J Phys Conf Ser. 2020;1691(1):12132.
36. Zohuri B, Mossavar-Rahmani F. Revolutionizing education: the dynamic synergy of personalized learning and artificial intelligence. Int J Adv Eng Manag Res. 2024;9(1):143–57.
37. Doo MY, Bonk CJ, Heo H. A meta-analysis of scaffolding effects in online learning in higher education. Int Rev Res Open Distrib Learn. 2020;21(3):60–80.
38. Chen H, Kumar P, Davis S. Emotional intelligence in education: the role of AI conversational systems. J Educ Psychol AI. 2023;12(4):45–67.
39. Gonzalez L, Patel R. Enhancing social skills through AI-driven collaborative learning. Int J Educ Res. 2022;78(2):102–15.
40. Mariyono D, Akmal NA, Annis NA. Navigating ethical dilemmas in AI-enhanced language education: addressing bias and ensuring inclusivity. J Educ Technol Stud. 2024;34(2):45–67.
41. Siddals S, Torous J, Coxon A. It happened to be the perfect thing: experiences of generative AI chatbots for mental health. NPJ Mental Health Res. 2024;3(1):48.
42. Rane NL, Choudhary SP, Rane J. Education 4.0 and 5.0: integrating artificial intelligence for personalized and adaptive learning. J Educ AI Integr. 2024;21(3):245–89.
43. Benitta G, Jimsy AH. An integrated (Hybrid-Based) besuited education system design for better models. J Educ AI Syst. 2024;27(1):123–47.
44. Ghimire A, Edwards J. Generative AI adoption in the classroom: TAM and IDT analysis. AI in Educ J. 2024;41(3):245–67.
45. Mitra A. AgentInstruct: toward generative teaching with agentic flows. In: *Proceedings of AI in Education*, 2024, pp. 120–143.
46. Liladhar NR. AI applications in automated assessments and learning. J Adapt Learn. 2024;39(1):89–104.
47. Dwivedi YK. So what if ChatGPT wrote it? multidisciplinary perspectives on opportunities, challenges, and implications of generative conversational AI for research, practice, and policy. Int J Inf Manage. 2023;71:102642.
48. Pandiaraj A. A comparative study on sentiment analysis using ML and DL. In: *Advanced Computing Conference Proceedings*, 2023, pp. 34–56.
49. Borovi F. Comparative analysis of generative AI tools in education. J Comp AI Stud. 2024;13(2):345–67.
50. Hooda M, Rana C, Dahiya O, Rizwan A, Hossain MS. Artificial intelligence for assessment and feedback to enhance student success in higher education. Math Probl Eng. 2022;2022:5215722.
51. Khosrawi-Rad B, Rinn H, Schlimbach R. Conversational Agents in Education? A Systematic Literature Review. In *Proceedings of the 30th European Conference on Information Systems (ECIS)*, Timi?oara, Romania, 2022.
52. Zhang K, Aslan AB. AI technologies for education: recent research & future directions. Comput Educ Artif Intell. 2021;2:100025.
53. Osakwe I, Chen G, Whitelock-Wainwright A. Towards automated content analysis of educational feedback: a multi-language study. Comput Educ Artif Intell. 2022;3:100059.
54. Shafiq DA, Marjani M, Habeeb RAA, Asirvatham D. Student retention using educational data mining and predictive analytics: a systematic literature review. IEEE Access. 2022;10:72480.
55. Kim J, Merrill K, Xu K, Sellnow DD. My teacher is a machine: understanding students. Perceptions of AI teaching assistants in online education. Int J Hum Comput Interact. 2020. https://doi.org/10.1080/10447318.2020.1801227.
56. Samala AD, et al. Unveiling the landscape of generative artificial intelligence in education: a comprehensive taxonomy of applications, challenges, and future prospects. Educ Inf Technol. 2024. https://doi.org/10.1007/s10639-024-12936-0.
57. Molenaar I. Towards hybrid human-AI learning technologies. Eur J Educ. 2022. https://doi.org/10.1111/ejed.12527.
58. Banihashem SK. The impacts of constructivist learning design and learning analytics on students engagement and self-regulation. Innov Educ Teach Int. 2021. https://doi.org/10.1080/14703297.2021.1890634.
59. Durlak J, Weissberg R, Dymnicki AB, Taylor RD, Schellinger KB. The impact of enhancing students social and emotional learning: a meta-analysis of school-based universal interventions. Child Dev. 2011;82(1):405–32.
60. Nascimento M et al. A learning management system accessible for visual, hearing and physical impairments. In *Universal Access in Human-Computer Interaction. Theory, Methods and Tools: 13th International Conference, UAHCI 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part I 21*, Springer, 2019, pp. 481–493.
61. Williams TG, et al. Integrating equity considerations into agent-based modeling: a conceptual framework and practical guidance. J Artif Soc Soc Simul. 2022. https://doi.org/10.18564/jasss.4816.
62. Lewthwaite S and Sloan D. Exploring pedagogical culture for accessibility education in computing science. In: *Proceedings of the 13th Web for All Conference*, 2016.
63. Bailey T, Morest VS. Defending the community college equity agenda. Baltimore: JHU Press; 2006.