UDC 004.8:378:51
DOI 10.56525/APMY4850

ARTIFICIAL INTELLIGENCE IN HIGHER EDUCATION MATHEMATICS: STATE-OF-THE-ART NARRATIVE REVIEW OF TEACHING, LEARNING, AND ASSESSMENT

Nuri Balta
SDU University, Almaty, Kazakhstan
e-mail: nuri.balta@sdu.edu.kz

Abstract. The rapid development of artificial intelligence (AI) systems is transforming educational practices across disciplines, including mathematics education in higher education. AI technologies, particularly generative AI and intelligent tutoring systems, introduce new opportunities for supporting teaching, learning, and assessment while simultaneously raising important pedagogical, ethical, and methodological challenges. This paper presents a state-of-the-art narrative review of recent research on AI in mathematics education in higher education. The review synthesizes literature from the past five to seven years and organizes it around three interconnected domains: teaching mathematics with AI, learning mathematics with AI, and assessment with AI. The analysis shows that AI has the potential to support instructional planning, adaptive feedback, and interactive explanations, thereby expanding possibilities for differentiated learning and formative assessment. At the same time, the literature shows critical concerns related to conceptual understanding, overreliance on automated solutions, assessment validity, authorship, and academic integrity. Particular attention is given to the design of AI-aware tasks and the need to foreground reasoning, interpretation, and mathematical communication in assessment practices. The review identifies key research gaps concerning classroom integration of AI, students' interaction with AI as a learning partner, and the development of AI-resilient assessment strategies. Implications are discussed for teacher education, curriculum design, and educational policy.

Keywords: Artificial intelligence in education; generative AI; mathematics education; higher education; AI-supported learning; AI-aware assessment; teacher competencies; AI literacy.

Introduction

The increasing availability of artificial intelligence (AI) based systems has begun to influence educational research and practice across subject areas, including mathematics education. Recent publications emphasize that AI systems differ from earlier generations of digital technologies in that they are capable of adapting to user input, generating responses, and supporting decision-making processes in instructional and learning contexts (1. Holmes et al., 2019) [1]. As a result, AI has been discussed as an additional instructional tool, and as a technological development with the potential to reshape established educational practices.

In mathematics education, digital technologies have long been studied with respect to their roles in teaching, learning, and assessment. Research has documented how tools such as computer algebra systems (CAS), dynamic geometry environments, and multimedia resources can support mathematical activity, while also highlighting the challenges teachers face when integrating such technologies into classroom practice (Clark-Wilson et al., 2021) [2]. More recently, policy and research reports have suggested that AI systems introduce new questions that extend beyond those previously associated with digital technologies, particularly because AI systems can generate mathematical explanations, provide adaptive feedback, and interact with learners in real time (Shah, 2023; UNESCO, 2023) [3, 4].

These developments have created a sense of urgency within mathematics education. Mathematics plays a central role in AI-related societal developments, both as a foundational discipline for technological innovation and as a school subject concerned with reasoning, modeling, and problem solving [3]. At the same time, [4] has cautioned that the increasing use of generative AI in education raises immediate concerns related to equity, transparency, and assessment practices, all of which are highly relevant for mathematics classrooms.

Against this background, several guiding questions emerge. First, what changes does AI bring for teachers of mathematics in higher education? Research has begun to examine how AI systems may influence teachers' professional roles, particularly with respect to instructional decision-making, responsibility, and pedagogical judgment (Kasneci et al., 2023; Selwyn, 2019) [5, 6]. These authors emphasize that AI should be understood as shaping instructor work rather than replacing it, while also noting the need for instructors to critically evaluate AI-generated outputs.

Second, how does AI mediate students' learning of mathematics in higher education? Earlier work on intelligent tutoring systems conceptualized AI as a form of classroom support capable of guiding learners through feedback and adaptive responses (du Boulay, 2016) [7]. More recent frameworks distinguish between different ways in which AI may be involved in learning processes, ranging from AI-directed instruction to AI-supported and AI-empowered learning environments (Ouyang et al., 2021) [8]. These perspectives shows that AI mediates learning by influencing how students engage with tasks, representations, and feedback, rather than by simply delivering content.

Third, what happens to assessment validity in the presence of AI in higher education? Educational measurement scholars have argued that AI-generated responses challenge traditional assumptions about authorship, construct representation, and the interpretation of student performance (Dorsey & Michaels, 2022; Williamson & Ecker, 2020) [9, 10]. These concerns are particularly salient in mathematics education, where assessment has traditionally relied on written solutions as evidence of students' reasoning and understanding.

The purpose of this study is to provide a state-of-the-art overview of research addressing these issues in mathematics in higher education. Following the structure adopted in recent work in the field (Weigand et al., 2024) [11], we focus on three interrelated areas: teaching mathematics with AI, learning mathematics with AI, and assessment with AI. By synthesizing recent research across these areas, this paper aims to clarify current conceptualizations, identify emerging research orientations, and outline challenges that require further investigation

Terminology and Methodological Considerations

As in the broader literature on digital technologies, research on AI in education employs a variety of partially overlapping terms. These terms are sometimes used interchangeably, despite referring to conceptually distinct systems and practices. To avoid ambiguity, we briefly clarify the terminology adopted in this paper.

AI is used here as an overarching term to refer to computer systems designed to perform tasks that typically require human intelligence, such as reasoning, learning from data, problem solving, and decision making. In educational contexts, AI is commonly associated with systems that can adapt to user input, analyze learner data, or generate responses based on computational models.

Within this broad category, generative AI refers to AI systems based on large-scale models that are capable of producing new content, such as text, code, images, or mathematical solutions, in response to user prompts (Bommasani et al., 2021) [12]. Recent educational research has emphasized that generative AI differs from earlier educational technologies because it can autonomously generate explanations and solution paths rather than merely supporting predefined instructional sequences [5].

A longer-standing line of research concerns intelligent tutoring systems (ITS). These systems are designed to model aspects of students' knowledge or problem-solving processes and to provide feedback, hints, or instructional guidance accordingly [7]. ITS typically rely on structured domain models and learner models and are therefore conceptually distinct from generative AI systems, which do not necessarily incorporate explicit representations of curricular goals or student understanding.

Related to ITS are adaptive learning systems, which personalize instructional content or learning trajectories based on learner data. Such systems adjust task sequences, feedback, or difficulty levels through predefined rules or learner models, often with the aim of supporting individualized learning pathways. While adaptive systems may incorporate AI techniques, they do not necessarily generate new content in the way generative AI systems do.

Another relevant concept is learning analytics, which refers to the measurement, collection, analysis, and reporting of data about learners and their contexts for the purpose of understanding and optimizing learning and learning environments. Learning analytics focuses primarily on data-driven insight and decision support rather than on direct instructional interaction with learners and it is therefore analytically distinct from tutoring or generative systems.

Finally, recent discussions have emphasized the importance of distinguishing between AI-assisted and AI-generated work. AI-assisted work refers to student output in which AI tools support human authorship, for example by offering feedback or suggestions, whereas AI-generated work refers to outputs primarily produced by AI systems themselves [4]. This distinction has become particularly important in relation to assessment, as AI-generated responses challenge traditional assumptions about authorship and construct validity in educational measurement [9].

In this paper, we use artificial intelligence as an umbrella term, while distinguishing between generative AI, intelligent tutoring systems, adaptive learning systems, and learning analytics where analytically relevant. These distinctions provide a conceptual basis for the subsequent synthesis of research on teaching, learning, and assessment in mathematics education.

Methodological Positioning of the Study

This paper adopts a state-of-the-art approach, also referred to as a narrative review, with the aim of synthesizing and conceptually organizing recent research on AI in mathematics education in higher education. State-of-the-art reviews are characterized by an expert-driven synthesis of influential and emerging work, with a primary focus on identifying dominant research orientations, conceptual developments, and open questions within a field (Barry et al., 2022) [13]. Unlike systematic reviews, such reviews do not aim at exhaustive coverage or replicable search procedures but instead seek to provide an integrative perspective that reflects the current status of research.

Consistent with this methodological orientation, the present review draws on a broad range of sources, including peer-reviewed journal articles, conference proceedings, and selected policy and research reports. Previous methodological analyses have emphasized that narrative and state-of-the-art reviews often rely on multiple source types in order to capture both established research findings and emerging discussions that may not yet be fully represented in journal publications [11]. Conference contributions and policy documents are therefore included where they provide insight into current trends, conceptual debates, or research directions relevant to mathematics education. Literature was identified through searches in databases such as Scopus, Web of Science, and Google Scholar, using keywords related to artificial intelligence, mathematics education, and higher education.

The time frame of the literature considered is primarily the most recent five to seven years, reflecting the rapid development of AI technologies and their educational applications. At the same time, earlier seminal contributions are included selectively where they provide necessary conceptual or historical grounding for understanding contemporary research (Barry et al., 2022) [14]. This combination of recent and foundational work is typical of state-of-the-art reviews, which aim to situate current developments within a broader research trajectory.

Methodologically, the review is non-systematic and expert-driven. Narrative reviews have been shown to play a complementary role to systematic and scoping reviews by enabling theoretical integration, critical interpretation, and conceptual clarification that extend beyond procedural mapping of the literature. While systematic and scoping reviews prioritize transparency and reproducibility of search and screening processe), the present approach emphasizes analytical coherence and interpretive synthesis.

The purpose of this methodological choice is field mapping and agenda setting. By organizing the literature around the interrelated themes of teaching mathematics with AI, learning mathematics with AI, and assessment with AI, the survey seeks to clarify how current research conceptualizes the role of AI in mathematics education, to identify converging and diverging perspectives, and to highlight issues that warrant further empirical and theoretical investigation [11, 13].

Teaching Mathematics with AI

The increasing presence of AI in educational contexts has prompted renewed discussion about the competencies required of teachers, including instructors of mathematics in higher education. While earlier research on digital competence focused on teachers' ability to integrate technological tools into instruction, recent literature emphasizes that AI introduces additional demands related to understanding system behavior, exercising pedagogical judgment, and assuming ethical and epistemic responsibility [3,14]. As a result, teachers' competencies in AI-rich environments cannot be reduced to technical proficiency alone.

A central component of teachers' AI competencies is AI literacy. AI literacy has been defined as a set of understandings and skills that enable individuals to critically engage with AI systems, including knowledge of how such systems function, what their limitations are, and how their outputs should be interpreted (Guvercin, 2025; Long & Magerko, 2020) [15,16]. Reviews of the educational literature further emphasize that AI literacy for teachers involves awareness of data use, algorithmic bias, and uncertainty in AI-generated responses. For mathematics teachers, AI literacy is particularly relevant because AI systems may produce mathematically correct-looking solutions or explanations whose underlying reasoning is opaque or incomplete.

Beyond technical understanding, the literature consistently shows the importance of pedagogical judgment in AI-supported teaching. Several authors caution against framing AI as a substitute for teachers' instructional decision-making, instead indicating that AI systems should be understood as tools whose outputs require professional interpretation and contextualization [6]. Studies examining generative AI in educational settings similarly argue that teachers must decide when, how, and for what purposes AI-generated content is pedagogically appropriate, rather than relying on automated responses uncritically [5]. In mathematics education, this judgment is closely connected to decisions about task selection, explanation quality, and the balance between procedural fluency and conceptual understanding.

Closely related to pedagogical judgment is teachers' ethical and epistemic responsibility when working with AI systems. Policy-oriented and scholarly discussions underline that educators remain accountable for instructional decisions, even when AI tools are involved in generating content or feedback [4]. From an epistemic perspective, concerns have been raised about the trustworthiness, transparency, and explainability of AI-generated outputs, particularly in domains such as mathematics where correctness and justification are central [10]. Ethical frameworks for AI further stress that teachers need to be attentive to issues of fairness, bias, and the responsible use of learner data.

Rather than proposing entirely new competence models, recent research suggests that teachers' AI competencies can be understood as extensions of existing frameworks, such as Technological Pedagogical Content Knowledge (TPACK), Artificial Intelligence Pedagogical Content Knowledge (AIPACK) and DigCompEdu. The TPACK framework conceptualizes teacher knowledge as an integration of content, pedagogy, and technology, and has been widely used to analyze teachers' technology-related competencies in mathematics education. The AIPACK framework (Balta, 2025) [17] conceptualizes teacher knowledge as an integration of content, pedagogy, and AI. Similarly, DigCompEdu provides a structured description of educators' digital competences, including professional engagement, pedagogical use of technology, and assessment practices. Recent reviews argue that these frameworks remain relevant in the context of AI, but require refinement to account for features such as automation, adaptivity, and content generation.

Teachers as Designers in AI-supported Environments

In mathematics education, teachers have long been conceptualized as implementers of curriculum materials and also as designers of tasks and learning environments. Research on digital technologies has emphasized that teachers' design work involves selecting, adapting, and sequencing tasks, as well as orchestrating tools and representations in ways that support students' mathematical activity. The introduction of AI systems extends this design role by introducing new forms of interaction between teachers, tasks, and technological resources.

One emerging aspect of teachers' design work in AI-supported environments concerns prompt design as task design. Studies on generative AI in education show that the prompts provided to AI systems strongly influence the nature, quality, and pedagogical relevance of the generated responses [5]. From an educational perspective, prompts can therefore be understood as instructional artifacts that encode learning goals, constraints, and expectations, rather than as purely technical inputs. While research on prompt design has largely developed outside mathematics education, it aligns with earlier findings that task formulation plays a central role in shaping students' engagement and reasoning processes.

Closely related is the challenge of designing AI-supported mathematical tasks. Recent discussions in mathematics education suggest that AI systems may be used to support task variation, generate examples or counterexamples, and provide alternative solution strategies, but that these possibilities need to be embedded within coherent task designs that foreground mathematical meaning (Zreik, 2024) [18]. From this perspective, AI does not replace established principles of task design; rather, it introduces additional design parameters related to automation, adaptivity, and content generation. Earlier theoretical work on instrumentation and teachers' documentation work remains relevant here, as it emphasizes that tools shape mathematical activity through their affordances and constraints, and that teachers play a key role in aligning these affordances with instructional intentions.

A further dimension of teachers' design work involves the orchestration of human–AI interaction in classrooms. Research on technology-rich mathematics classrooms has shown that teachers actively manage how tools are introduced, when they are used, and how students' interactions with these tools are discussed and reflected upon. This notion of instrumental orchestration provides a useful lens for understanding AI-supported instruction, as it foregrounds the teacher's responsibility for structuring learning situations and mediating students' engagement with technological resources. In the context of AI, orchestration includes decisions about when AI systems are appropriate, how their outputs are interpreted, and how AI-mediated activity is connected to mathematical discourse.

Policy-oriented and critical perspectives further stress that AI systems should be embedded within teacher-led instructional frameworks, rather than operating autonomously in classrooms [1, 6]. These authors emphasize that teachers remain accountable for pedagogical decisions and learning outcomes, even when AI systems contribute to content generation or feedback. As a result, human–AI orchestration in mathematics classrooms is not primarily a technical issue, but a pedagogical one, involving judgment about instructional goals, equity, and the role of automation in learning.

AI-enhanced Teaching Practices

Recent literature suggests that AI may influence teaching practices in mathematics education in multiple ways, particularly with respect to lesson planning, differentiation, and instructional support. Research and policy reports emphasize that AI systems can assist teachers by providing suggestions, resources, or alternative representations, while leaving pedagogical decision-making with the teacher. In this sense, AI is typically framed as a supportive technology rather than an autonomous instructional agent.

One area where AI has attracted attention is lesson planning. Studies on generative AI indicate that such systems can be used to generate draft lesson plans, examples, explanations, or assessment items that teachers may adapt to their instructional goals [5]. These contributions are described as potentially reducing preparation time and supporting idea generation, particularly when teachers retain control over task selection, sequencing, and adaptation. However, existing research also stresses that AI-generated plans require careful review, as they may not align with curricular intentions or pedagogical principles without teacher intervention [1].

AI systems have also been discussed in relation to differentiation and scaffolding. Earlier work on intelligent tutoring systems and adaptive learning environments demonstrates that AI-based systems can provide individualized feedback, hints, or task progression based on learner responses [7]. These features have been associated with opportunities for supporting learners at different levels of understanding and pacing instruction more flexibly. At the same time, the literature notes that adaptive support is most effective when integrated into teacher-led instructional designs, rather than functioning independently of classroom practices.

Alongside these opportunities, several authors show risks associated with AI-enhanced teaching practices. One concern relates to the potential deskilling of teachers if instructional decisions are increasingly delegated to automated systems [6]. From this perspective, reliance on AI-generated content or recommendations may reduce opportunities for teachers to engage in reflective planning and professional judgment. Related concerns involve overreliance on AI outputs, particularly when teachers or students accept generated explanations or solutions without critical evaluation [5].

Issues of bias and fairness further complicate the use of AI in teaching. Ethical analyses and policy guidance emphasize that AI systems may reflect biases present in their training data and that these biases can influence instructional content or feedback in subtle ways. In mathematics education, where correctness and justification are central, such biases raise questions about the trustworthiness of AI-generated explanations and examples. Scholars therefore stress that teachers remain responsible for monitoring AI use and for ensuring that instructional practices align with ethical and educational standards.

Learning Mathematics with AI

Recent literature increasingly conceptualizes AI as a tool used by learners, and as a co-participant in learning processes. This perspective reflects a shift from earlier views of educational technology as a passive medium toward understandings of learning as distributed across humans and technological systems. In the context of mathematics education, this shift raises questions about how AI systems participate in explanation, tutoring, and feedback, and how such participation reshapes students' engagement with mathematical ideas.

One line of research frames AI as an explainer, tutor, or feedback provider. Earlier studies on intelligent tutoring systems describe AI-based environments that model aspects of learners' knowledge and provide adaptive feedback or guidance during problem solving [7]; VanLehn, 2011). These systems are designed to respond to students' actions in ways that resemble certain features of human tutoring, such as step-by-step feedback or targeted hints. More recent discussions extend this view to generative AI systems, which can produce explanations or solution strategies in response to learners' prompts, thereby offering new forms of interaction between students and mathematical content.

Building on these developments, several authors propose understanding learning with AI through a humans-with-AI perspective, which emphasizes the joint activity of learners and intelligent systems. This perspective resonates with earlier theoretical approaches in mathematics education that conceptualized learning as mediated by tools and representations, while acknowledging that AI systems differ from earlier technologies due to their adaptive and generative capabilities. Frameworks distinguishing between AI-directed, AI-supported, and AI-empowered learning further indicate that AI may play different roles in learning activity, ranging from guiding learners' actions to supporting reflection and exploration [8]. In mathematics education, such distinctions have been used to discuss how AI systems may influence students' reasoning processes and engagement with representations [18].

At the same time, the literature draws attention to tensions between productive struggle and the availability of immediate answers provided by AI systems. Research on learning emphasizes that struggling with challenging problems can play a productive role in increasing conceptual understanding and deeper learning. From this perspective, the capacity of AI systems to generate complete solutions or explanations on demand may risk short-circuiting opportunities for sense-making if not carefully mediated. Studies examining generative AI in educational contexts caution that learners may become overly reliant on AI-generated responses, potentially reducing engagement in mathematical reasoning and problem solving (Domínguez-González et al., 2023; Kasneci et al., 2023) [5, 19].

Multiple resources: AI, CAS, DGS, and Paper-and-Pencil

Research in higher mathematics education has long emphasized that learning typically involves the coordinated use of multiple resources, including symbolic tools, graphical representations, digital environments, and paper-and-pencil work. Rather than replacing existing resources, AI systems are increasingly discussed as becoming part of a broader ecology of tools with which students interact. From this perspective, understanding learning with AI requires attention to how learners transition within, across, and beyond different resources.

Theoretical work on instrumentation shows that learners' engagement with mathematical tools involves processes through which artifacts become instruments for thinking, shaped by both the affordances of the tools and the users' schemes of action Subsequent research has shown that students frequently move between different tools, such as CAS, dynamic geometry software (DGS), and paper-and-pencil, during mathematical activity, and that these transitions play a central role in meaning making. Recent discussions suggest that AI systems can be understood within this framework as additional resources that learners may draw upon, rather than as standalone environments [18].

Within AI-supported environments, transitions within tools may occur when learners interact iteratively with an AI system, for example by refining prompts or interpreting feedback. Transitions across tools involve movement between AI systems and other resources, such as using AI-generated explanations alongside CAS output or DGS constructions. Finally, transitions beyond tools refer to moments when learners disengage from digital environments to reason with paper-and-pencil or verbal explanations. Prior research on technology integration indicates that such transitions are not incidental but are integral to students' mathematical activity, as they support comparison, verification, and reflection.

A related line of research concerns representational fluency, understood as learners' ability to interpret, transform, and coordinate multiple representations of mathematical ideas. Conceptual frameworks emphasize that learning is supported when learners can flexibly move between symbolic, graphical, numerical, and verbal representations, while recognizing their distinct roles and limitations. Technology-enhanced learning environments have been shown to influence how students engage with representations, sometimes foregrounding particular forms while backgrounding others.

In AI-supported contexts, representational fluency takes on additional dimensions. AI systems can generate representations dynamically, including symbolic expressions, graphs, verbal explanations, or visualizations, potentially supporting learners' exploration of mathematical relationships. At the same time, the literature cautions that automatically generated representations may obscure underlying structures if learners engage with them uncritically. Studies comparing different technology-enhanced environments suggest that learning outcomes depend on the availability of representations, and on how learners are guided to interpret and connect them (Ng et al., 2020) [20].

Opportunities and Risks for Mathematical Understanding

The growing use of AI in mathematics learning environments has prompted renewed discussion about its potential contributions to, and risks for, students' mathematical understanding. Existing research shows that AI systems may support learning in certain respects, while also introducing challenges related to misconceptions, reliability of explanations, and equity. From a mathematics education in higher education perspective, these opportunities and risks need to be examined in relation to conceptual understanding, the emergence of erroneous reasoning, and differential access to AI-supported learning.

With respect to conceptual understanding, research on technology use in mathematics education consistently emphasizes that digital tools can support meaning-making when they facilitate connections between representations and encourage reflection on underlying concepts. Recent discussions of AI in mathematics education extend this view by showing that AI systems may offer new forms of explanation, feedback, and representational support that can potentially enrich students' engagement with mathematical ideas [18]. However, studies of advanced tools such as computer algebra systems also caution that conceptual understanding does not automatically follow from tool use; rather, it depends on how learners interpret outputs and justify results within mathematically meaningful activity.

At the same time, the literature points out to the risks related to misconceptions and incorrect explanations, which have gained renewed attention in discussions of generative AI. Research on learning has long shown that learners' misconceptions can be locally coherent and resistant to change, making them difficult to detect and address. In the context of AI-supported learning, this concern is compounded by the fact that AI systems may generate explanations or solutions that appear plausible while being mathematically incorrect. Studies examining large language models document systematic instances of such erroneous outputs, often referred to as "hallucinations," which may mislead learners if accepted uncritically (Borji, 2023) [20]. Educational analyses further warn that the fluency and confidence of AI-generated responses can obscure underlying errors, posing challenges for learners who lack the expertise to evaluate correctness [5].

Beyond issues of correctness, scholars also raise concerns about equity and access in AI-supported mathematics learning. Critical analyses of AI in education argue that access to AI technologies is unevenly distributed and that existing social and educational inequalities may be reinforced rather than mitigated by their use [6]. Policy guidance similarly emphasizes that AI systems may embed biases related to language, culture, or prior data, which can affect the quality and fairness of learning opportunities [4]. From this perspective, AI-supported learning environments may differentially benefit students depending on their access to technology, prior knowledge, and support structures. Broader analyses of AI in education further situate these concerns within political and economic contexts that shape how AI systems are developed and deployed, showing the need for transparency and accountability Williamson & Eynon, 2020) [21].

Assessment with AI

AI in Formative and Summative Assessment

The increasing use of AI in higher educational contexts has prompted renewed discussion about its role in assessment, particularly with respect to formative and summative purposes. Existing research distinguishes between using AI with assessment, where AI supports assessment processes, and assessment conducted through AI, where AI systems assume a more central role in generating judgments or scores.

Several authors argue that AI can contribute to assessment practices when it is used in support of human judgment, rather than as a substitute for it. Analyses of AI and automated scoring show that AI systems may assist with tasks such as organizing evidence, generating feedback, or supporting scoring processes, but that assessment decisions ultimately require interpretive judgment to ensure construct validity and accountability. From this perspective, assessment with AI refers to the use of AI as a tool that augments assessment practices without displacing human responsibility.

In contrast, assessment through AI involves a greater delegation of evaluative functions to automated systems. Research in educational measurement cautions that such delegation raises concerns related to transparency, interpretability, and the alignment between assessment outcomes and intended constructs [9]. Earlier work on digital assessment similarly stresses the importance of distinguishing between technology-supported assessment and technology-driven assessment, particularly in contexts where assessment outcomes carry significant consequences for learners. These concerns are especially salient in mathematics education, where assessment often serves as evidence of reasoning, understanding, and justification.

A further line of research shows the potential of AI to support continuous feedback systems, which are closely associated with formative assessment. Studies on formative feedback emphasize that timely, specific, and task-focused feedback can support learning when it is integrated into instructional activity. AI-based systems, including intelligent tutoring systems, have been shown to provide ongoing feedback during problem solving, adapting responses to learners' actions and offering guidance without assigning final judgments [7]. Such feedback mechanisms align with conceptions of formative assessment that emphasize learning support rather than evaluation.

The literature consistently indicates that AI aligns more readily with formative purposes, where feedback and adaptation are central, than with summative purposes that require stable interpretations and defensible judgments. Policy-oriented and conceptual analyses suggest that AI-supported feedback systems may contribute to formative assessment by enabling more frequent feedback cycles and supporting teachers in monitoring student progress [1]. However, these contributions do not change the fundamental distinction between formative and summative assessment. Formative assessment is defined by its function in supporting learning, whereas summative assessment serves to evaluate learning outcomes at particular points in time.

AI-generated Feedback and Adaptive Assessment

A growing body of literature discusses the use of AI to generate feedback and to support adaptive assessment processes. These developments raise questions about the reliability and transparency of AI-generated feedback, as well as about the balance between personalization and standardization in assessment practices. From an assessment perspective, these issues are central to determining how AI-supported systems align with established principles of validity and fairness.

Without insight into how feedback is generated or how learner responses are interpreted, it becomes difficult for educators and learners to judge the appropriateness of AI-supported assessment outcomes. Research on AI-based and automated assessment show that AI systems can produce feedback and scoring with a high degree of consistency, particularly when models are trained on large datasets and apply stable decision rules (Williamson & Bejar, 2020). Such consistency is often presented as an advantage in comparison to human scoring, which may vary across raters or contexts. However, assessment scholars caution that reliability alone is insufficient; AI-generated feedback must also be transparent and interpretable in order to support valid assessment use [9].

Concerns about transparency are particularly salient in adaptive assessment systems, where feedback and task selection are dynamically adjusted based on inferred learner states. Research on adaptive learning technologies indicates that such systems rely on underlying models of learner knowledge, which necessarily involve assumptions about learning progressions and performance indicators. While these models can support responsive feedback, the literature emphasizes that their functioning should be made explicit to ensure that teachers can interpret results and intervene appropriately. From this perspective, transparency is closely linked to accountability, as opaque systems limit opportunities for professional judgment and critical evaluation.

A further theme in the literature concerns the tension between personalization and standardization. Adaptive assessment systems are designed to tailor feedback, task difficulty, or learning paths to individual learners, thereby supporting differentiated learning experiences. Such personalization aligns well with formative assessment purposes, where variation across learners is both expected and pedagogically meaningful. At the same time, assessment theory show that standardization is essential when assessment outcomes are used for summative or comparative purposes.

Studies of adaptive educational systems therefore stress the importance of clarifying assessment purposes when deploying AI-generated feedback. Personalized feedback may be appropriate and beneficial for supporting learning, but its use in evaluative contexts requires careful consideration of how results are interpreted and compared. Assessment researchers caution that blending formative and summative purposes within the same adaptive system can lead to ambiguity about what assessment outcomes represent and how they should be used [9].

Validity, Authorship, and Academic integrity

The integration of AI into educational contexts raises fundamental questions about assessment validity, authorship, and academic integrity. In particular, the availability of AI systems that can generate text, solutions, or explanations challenges traditional assumptions about what constitutes student work and how assessment evidence should be interpreted. From an assessment perspective, these issues are central because validity depends on a clear relationship between observed performance and the constructs being assessed.

Assessment scholars argue that AI complicates the interpretation of assessment evidence by weakening the link between a student's knowledge or skills and the products submitted for evaluation [9]. When AI systems contribute to the generation of responses, it becomes less clear whether assessment outcomes reflect individual understanding, collaborative activity, or automated assistance. Similar concerns have been raised in discussions of automated scoring, where the involvement of technological systems requires explicit consideration of how performance evidence is produced and interpreted [10].

Related discussions focus on authorship and originality. Research on AI-assisted writing and problem-solving shows that AI blurs established distinctions between independent work, permitted assistance, and inappropriate substitution. Rather than framing these challenges solely in terms of misconduct, recent scholarship emphasizes the need to clarify expectations about disclosure, collaboration, and the acceptable use of AI tools. From this perspective, academic integrity concerns are closely connected to assessment design and institutional norms, rather than to individual student behavior alone.

These challenges have prompted renewed interest in authentic assessment redesign. The literature on authentic assessment emphasizes tasks that require learners to demonstrate understanding through meaningful performances, contextualized problem solving, and reflective processes. Such approaches shift the focus from static products to processes and reasoning, thereby reducing the relevance of simple reproduction or outsourcing of answers. In AI-rich contexts, authentic assessment is discussed as a way to preserve validity by aligning assessment tasks more closely with learning goals that cannot be easily delegated to automated systems.

Policy-oriented analyses and integrity research further argue that responses to AI-related challenges should extend beyond detection or prohibition. Instead, they call for systemic approaches that integrate transparency, ethical guidance, and pedagogical redesign. Critical perspectives caution that surveillance-oriented responses may raise equity and trust concerns, while failing to address underlying questions about what assessment is intended to measure [6]. From this viewpoint, academic integrity in the age of AI is a technical issue, and a conceptual and educational one.

Task Design for AI-aware Assessment

The availability of AI tools capable of generating complete solutions and explanations has intensified attention to assessment task design. In AI-rich contexts, the validity of assessment depends increasingly on tasks that elicit evidence of students' reasoning, interpretation, and decision-making, rather than on final products alone. As a result, existing research on authentic and process-oriented assessment has gained renewed relevance.

One line of work emphasizes the value of oral examinations, portfolios, and process evidence as sources of assessment information. Research on educative and authentic assessment argues that understanding is more accurately inferred when learners are required to explain their reasoning, justify decisions, and reflect on their work. Oral questioning and dialogic assessment practices allow assessors to probe students' thinking and to clarify how responses are constructed, thereby reducing ambiguity about authorship and assistance. Similarly, portfolio-based assessment foregrounds learning processes over time by documenting intermediate products, revisions, and reflections, offering a richer basis for interpretation than single, decontextualized responses.

The assessment literature further suggests that process-oriented evidence aligns closely with formative assessment principles. Classroom-based assessment research highlights that evidence gathered through discussion, observation, and ongoing documentation can support valid inferences about learning while remaining integrated with instruction. In AI-supported environments, such approaches may provide insight into how learners engage with tasks and resources, including AI systems, rather than focusing solely on outputs that could have been externally made.

A complementary body of research addresses the design of AI-resilient mathematics tasks, that is, tasks whose assessment value does not depend on restricting access to tools but on the nature of the mathematical activity involved. Tasks that foreground creative reasoning provide assessment evidence that extends beyond what can be captured by automated solution generation. Studies on task design in mathematics education emphasize that tasks requiring explanation, comparison of strategies, and interpretation of errors are less amenable to simple answer reproduction (Swan, 2005). Similarly, theoretical work on mathematical reasoning distinguishes between imitative reasoning, which follows known procedures, and creative reasoning, which requires learners to construct and justify solution paths (Lithner, 2008).

Research on cognitively demanding tasks further supports this perspective. Tasks that maintain high levels of cognitive demand require learners to make sense of concepts, connect representations, and articulate reasoning throughout the problem-solving process. In the context of AI, such tasks shift the focus of assessment from producing correct answers to demonstrating understanding through argumentation and interpretation. Assessment scholars note that preserving validity under these conditions depends on aligning task demands with the intended constructs, rather than on attempting to control tool use directly [9].

From a mathematics education standpoint, competence-oriented frameworks reinforce the importance of assessing processes such as reasoning, communication, and modeling (Niss & Højgaard, 2019). Recent discussions of AI in mathematics education argue that the ease with which solutions can be generated heightens the importance of tasks that assess how learners interpret results, evaluate methods, and connect representations [18]. These perspectives show continuity with longstanding principles of task design, while also indicating their renewed significance in AI-aware assessment contexts.

Conclusion and Future Agenda

This review set out to examine how AI is reshaping mathematics education across the interrelated domains of teaching, learning, and assessment. Rather than treating AI as a discrete innovation, the review situated AI within established theoretical and methodological traditions of mathematics in higher education. From this perspective, AI emerges not as a disruptive replacement of existing practices but as a development that amplifies longstanding questions about pedagogy, learning processes, and assessment validity.

Across the literature, several promising directions can be identified. In teaching, AI appears to offer support for lesson planning, task adaptation, and feedback generation, particularly when teachers retain professional judgment and design responsibility. The extension of existing competence frameworks, such as TPACK, AIPACK and DigCompEdu, suggests continuity rather than rupture in how teacher competencies are conceptualized.

In learning, the framing of AI as a co-participant highlights opportunities for interactive explanation, adaptive feedback, and representational support. When integrated alongside other mathematical resources, such as CAS, DGS, and paper-and-pencil, AI may contribute to learners' engagement with multiple representations and to exploratory activity. Similarly, in assessment, AI shows potential in supporting formative practices through continuous feedback systems that align with established theories of formative assessment.

Unresolved Issues. At the same time, the literature reveals a number of unresolved issues. A persistent concern relates to the balance between automation and professional judgment. While AI systems can generate explanations, solutions, and feedback, questions remain about when such outputs support learning and when they risk undermining conceptual understanding or productive struggle.

In assessment, unresolved tensions persist between personalization and standardization, particularly when adaptive systems are used in contexts that require comparability and defensible judgments. The meaning of assessment evidence becomes increasingly ambiguous when AI contributes to the production of student work, raising ongoing questions about validity, authorship, and integrity.

Equity and ethical considerations also remain insufficiently resolved. Differences in access to AI tools, variability in data quality, and the potential for algorithmic bias complicate claims about the educational benefits of AI. These issues extend beyond technical design and point to broader structural and governance challenges.

Research Gap. The review points to several research gaps that warrant systematic investigation. Empirical research remains limited with respect to how teachers actually integrate AI into mathematics classrooms over time, including how they design tasks, orchestrate resources, and interpret AI-generated outputs. Similarly, more research is needed on how students engage with AI as a learning partner, particularly in relation to reasoning, sense-making, and the development of mathematical understanding. In assessment, there is a lack of empirical studies examining how AI-aware task designs function in practice and how different forms of process evidence are interpreted by teachers and institutions. Conceptual work is needed to refine theoretical frameworks that account for AI's adaptive and generative features while remaining grounded in established mathematics education theory.

Implications for Teacher Education. For teacher education, the findings suggest a need to move beyond technical training toward developing AI literacy that encompasses pedagogical judgment, ethical awareness, and assessment competence. Preparing mathematics instructors involves helping them critically evaluate AI outputs, design AI-supported tasks, and make informed decisions about when and how AI should be used.

In terms of curriculum design, AI shows the importance of tasks that foreground reasoning, interpretation, and representation. Curricula that emphasize mathematical processes rather than isolated products appear better aligned with AI-rich environments, where solution generation is readily available. Designing opportunities for students to engage critically with AI-generated representations and explanations becomes an increasingly relevant curricular concern.

At the level of policy and ethics, the literature shows the need for clear guidance on acceptable AI use, transparency, and accountability. Policies focused solely on detection or prohibition risk overlooking deeper questions about assessment purposes and educational values. Instead, policy frameworks need to address equity, data governance, and the ethical responsibilities of institutions, teachers, and technology providers.

Overall, the reviewed literature suggests that AI does not fundamentally alter the core aims of mathematics in higher education, but it does intensify existing challenges related to teaching quality, learning processes, and assessment validity. Addressing these challenges requires continued dialogue between mathematics education research, educational practice, and policy, as well as sustained empirical and theoretical work that keeps pedagogical goals at the forefront of AI integration.

REFERENCES

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

2. Clark-Wilson, A., Donevska-Todorova, A., Faggiano, E., Trgalová, J., & Weigand, H. G. (2021). *Mathematics education in the digital age*. Routledge.

3. Shah, P. (2023). *AI and the Future of Education: Teaching in the Age of Artificial Intelligence*. John Wiley & Sons.

4. UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

5. Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

6. Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.

7. du Boulay, B. (2016). Artificial intelligence as an effective classroom assistant. *IEEE Intelligent Systems, 31*(6), 76–81. https://doi.org/10.1109/MIS.2016.93

8. Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, *2*, 100020.

9. Dorsey, D. W., & Michaels, H. R. (2022). Validity arguments meet artificial intelligence in innovative educational assessment. *Journal of Educational Measurement*, *59*(3), 267-271.

10. Williamson, D. M., & Ecker, A. (2020). Automated scoring and the future of assessment. *Educational Measurement: Issues and Practice, 39*(3), 33–38. https://doi.org/10.1111/emip.12340

11. Weigand, H.-G., Trgalová, J., & Tabach, M. (2024). Mathematics teaching, learning, and assessment in the digital age. *ZDM–Mathematics Education, 56*(4), 525–541. https://doi.org/10.1007/s11858-024-01612-9

12. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., … Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv*. https://arxiv.org/abs/2108.07258

13. Barry, E. S., Merkebu, J., & Varpio, L. (2022). Understanding state-of-the-art literature reviews. *Journal of Graduate Medical Education*, *14*(6), 659-662.

14. Balta, N. (2023). Ethical Considerations in Using AI in Educational Research. *Journal of Research in Didactical Sciences, 2*(1), 14205. https://doi.org/10.51853/jorids/14205

15. Guvercin, S. (2025). The Urgent Case for AI Literacy in the 21st Century. International Educational Review, 3(1), 41-44. DOI: https://doi.org/10.58693/ier.313

16. Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). ACM. https://doi.org/10.1145/3313831.3376727

17. Balta, N. (2025). Artificial intelligence pedagogical content knowledge. *The European Educational Researcher, 8*(1), 1–3. https://doi.org/10.31757/euer.811

18. Zreik, M. (2024). Redefining mathematics education in the age of artificial intelligence: Challenges and opportunities. *Impacts of globalization and innovation in mathematics education*, 19-40.

19. Domínguez-González, M. A., Hervás-Gómez, C., Díaz-Noguera, M. D. & Reina-Parrado, M. (2023). Attention to Diversity from Artificial Intelligence. *The European Educational Researcher, 6*(3), 101-115. DOI: https://doi.org/10.31757/euer.633

20. Ng, O. L., Shi, L., & Ting, F. (2020). Exploring differences in primary students' geometry learning outcomes in two technology-enhanced environments: dynamic geometry and 3D printing. *International Journal of STEM Education*, *7*(1), 50.

21. Borji, A. (2023). A categorical archive of ChatGPT failures. *arXiv*. https://arxiv.org/abs/2302.03494

ВЫСШЕЕ МАТЕМАТИЧЕСКОЕ ОБРАЗОВАНИЕ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: СОВРЕМЕННЫЙ НАРРАТИВНЫЙ ОБЗОР ПРЕПОДАВАНИЯ, ОБУЧЕНИЯ И ОЦЕНИВАНИЯ

Нури Балта
Университет SDU, Алматы, Казахстан
e-mail: nuri.balta@sdu.edu.kz

Аннотация. Стремительное развитие систем искусственного интеллекта (ИИ) трансформирует образовательные практики в различных областях, включая математическое образование в высшей школе. Технологии ИИ, в частности генеративный ИИ и интеллектуальные обучающие системы, открывают новые возможности для поддержки преподавания, обучения и оценивания, одновременно создавая важные педагогические, этические и методологические вызовы. В данной статье представлен современный нарративный обзор последних исследований, посвящённых использованию ИИ в математическом образовании в высшей школе. Обзор обобщает литературу за последние пять–семь лет и структурирует её вокруг трёх взаимосвязанных направлений: преподавание математики с использованием ИИ, обучение математике с использованием ИИ и оценивание с применением ИИ.

Анализ показывает, что ИИ обладает потенциалом для поддержки планирования обучения, предоставления адаптивной обратной связи и интерактивных объяснений, тем самым расширяя возможности дифференцированного обучения и формирующего оценивания. В то же время в литературе отмечаются серьёзные проблемы, связанные с пониманием математических концепций, чрезмерной зависимостью от автоматизированных решений, валидностью оценивания, авторством и академической добросовестностью. Особое внимание уделяется разработке заданий с учётом использования ИИ и необходимости акцентировать внимание на рассуждении, интерпретации и математической коммуникации в практике оценивания.

В обзоре также выявляются ключевые пробелы в исследованиях, касающиеся интеграции ИИ в учебный процесс, взаимодействия студентов с ИИ как с партнёром в обучении, а также разработки устойчивых к ИИ стратегий оценивания. Обсуждаются выводы и рекомендации для подготовки преподавателей, разработки учебных программ и образовательной политики.

Ключевые слова: искусственный интеллект в образовании; генеративный ИИ; математическое образование; высшее образование; обучение с поддержкой ИИ; оценивание с учётом ИИ; компетенции преподавателя; ИИ-грамотность.

ЖОҒАРЫ БІЛІМДЕГІ МАТЕМАТИКАДА ЖАСАНДЫ ИНТЕЛЛЕКТ: ОҚЫТУ, ОҚУ ЖӘНЕ БАҒАЛАУДЫҢ ҚАЗІРГІ ЖАҒДАЙЫНА АРНАЛҒАН НАРРАТИВТІК ШОЛУ

Нури Балта
СДУ Университеті, Алматы, Қазақстан
e-mail: nuri.balta@sdu.edu.kz

Аңдатпа. Жасанды интеллект (ЖИ) жүйелерінің қарқынды дамуы түрлі салалардағы білім беру тәжірибелерін, соның ішінде жоғары білімдегі математикалық білім беруді де өзгертіп жатыр. ЖИ технологиялары, әсіресе генеративті ЖИ мен интеллектуалды оқыту жүйелері, оқыту, оқу және бағалауды қолдаудың жаңа мүмкіндіктерін ұсына отырып, сонымен қатар маңызды педагогикалық, этикалық және әдіснамалық мәселелерді де туындатады. Бұл мақалада жоғары білімдегі математикалық білім беруде ЖИ қолдануға арналған соңғы зерттеулердің заманауи нарративтік шолуы ұсынылады. Шолу соңғы бес–жеті жылдағы әдебиеттерді жинақтап, оларды өзара байланысты үш бағыт бойынша жүйелейді: ЖИ көмегімен математиканы оқыту, ЖИ көмегімен математиканы оқу және ЖИ негізінде бағалау.

Талдау нәтижелері ЖИ-дің оқу процесін жоспарлауды қолдауға, бейімделген кері байланыс ұсынуға және интерактивті түсіндірулер беруге мүмкіндік беретінін, сол арқылы сараланған оқыту мен қалыптастырушы бағалаудың мүмкіндіктерін кеңейтетінін көрсетеді. Сонымен қатар, әдебиеттерде математикалық ұғымдарды түсіну, автоматтандырылған шешімдерге шамадан тыс тәуелділік, бағалаудың валидтілігі, авторлық және академиялық адалдыққа қатысты елеулі мәселелер де атап өтіледі. Ерекше назар ЖИ-ді ескеріп әзірленетін тапсырмаларды жобалауға және бағалау тәжірибесінде дәлелдеу, интерпретация және математикалық коммуникацияға басымдық берудің қажеттілігіне аударылады.

Шолуда сондай-ақ ЖИ-ді оқу процесіне енгізу, студенттердің ЖИ-мен оқу серіктесі ретінде өзара әрекеттесуі және ЖИ-ге төзімді бағалау стратегияларын әзірлеуге қатысты негізгі зерттеу олқылықтары анықталады. Педагогтарды даярлау, оқу бағдарламаларын әзірлеу және білім беру саясатына қатысты қорытындылар мен ұсыныстар талқыланады.

Түйін сөздер: білім берудегі жасанды интеллект; генеративті ЖИ; математикалық білім беру; жоғары білім; ЖИ қолдауымен оқыту; ЖИ-ге бейімделген бағалау; мұғалім құзыреттіліктері; ЖИ сауаттылығы.