International Journal of Instruction *January 2026 ● Vol.19, No.1*
e-ISSN: 1308-1470 ● www.e-iji.net *p-ISSN: 1694-609X*
pp. 257-280

Article submission code: 20250502225545
Received: 02/05/2025
Revision: 02/08/2025
Accepted: 10/08/2025
OnlineFirst: 03/10/2025

Can Artificial Intelligence (AI) Shape the Future of Teacher Education? Drawing Evidence-Based Insights for Teacher Preparation

Ronilo P. Antonio
College of Education, Bulacan State University, City of Malolos, Bulacan, Philippines, *ronilo.antonio@bulsu.edu.ph*

Louis Robert C. Sison
College of Education, Bulacan State University, City of Malolos, Bulacan, Philippines. *louisrobert.sison@bulsu.edu.ph*

As artificial intelligence (AI) reshapes education worldwide, its role in preservice teacher training has become increasingly significant yet underexplored. This meta-analysis synthesizes empirical evidence on the effectiveness of AI-based interventions in enhancing the learning outcomes of preservice teachers across cognitive, affective, and psychomotor domains. A systematic review of 5,880 studies led to the inclusion of 19 effect sizes from 11 rigorously selected studies spanning diverse geographic and disciplinary contexts. Using Hedges' g and a random-effects model to account for high heterogeneity, the analysis reveals a large overall effect size, affirming AI's substantial positive impact on teacher preparation. Moderator analyses highlight the influence of contextual factors such as country, learning outcome, and intervention duration. Studies from Palestine, Ethiopia, and the UAE, revealed strong effects which suggest the importance of local readiness and institutional support. Cognitive outcomes showed the most consistent gains, while affective and psychomotor results were more variable. Most interventions operated at the Augmentation level of the SAMR model, a framework that categorizes levels of technology integration in education, indicating that transformative uses of AI in teacher education remain limited. This study offers critical insights for policymakers, educators, and global teacher education programs aiming to integrate AI effectively. It underscores the need for context-sensitive, ethically responsible, and pedagogically innovative AI applications that go beyond enhancement toward true educational transformation. Future research should address geographic gaps, explore under-examined learning outcomes, and promote equitable access to AI technologies to ensure inclusive and sustainable impact in teacher education systems worldwide.

Keywords: artificial intelligence, emerging technologies, teacher education, meta-analysis, AI, teacher training

Citation: Antonio, R., P., & Sison, L. R. C. (2026). Can artificial intelligence (AI) shape the future of teacher education? Drawing evidence-based insights for teacher preparation. *International Journal of Instruction*, *19*(1), 257-280.

INTRODUCTION

The rise of technologies embedding Artificial Intelligence (AI) has enabled the development of sophisticated educational tools to support the teaching-learning process, including intelligent tutoring systems and automated grading systems. The harnessing of AI is transformational within educational settings, where educators can seamlessly personalize instruction, develop efficient lesson plans and instructional materials, monitor learner achievement in real time, and produce multiple forms of assessments, thereby creating a more responsive learning environment (Celik et al., 2022; Hu et al., 2022; Zawacki-Richter et al., 2019). The capacity of AI to positively influence student engagement, equity, and performance across varying educational circumstances is widely recognized as one of its most notable benefits (Cáceres-Nakiche et al., 2024; Hu et al., 2022).

In the present education parlance, AI is now being integrated into teacher education, which is considered a critical stage in a teacher's early professional life, particularly the pre-service teaching journey. The use of AI in their training is increasingly emphasized as teacher education programs aim to prepare candidates for the demands of 21st-century teaching (Celik et al., 2022; Hu et al., 2022; Zawacki-Richter et al., 2019). Currently, AI technologies are supporting preservice teachers with lesson planning, teaching practice simulations, pedagogical content knowledge acquisition, and self-reflection activities (Celik et al., 2022; Hu et al., 2022; Zawacki-Richter et al., 2019). The integration of AI into teacher education programs offers possibilities for improved learning outcomes, individualized instruction, tailored learning experiences, and the development of critical and flexible thinking among preservice teachers (Cáceres-Nakiche et al., 2024; Hu et al., 2022; Meylani, 2024).

Despite the immense potential of AI in pre-service teacher education, critical challenges remain. These include concerns about inappropriate use of technology, data privacy, ethical use, and the preparedness of both institutions and teachers to embed these tools in instruction (Celik et al., 2022; Hu et al., 2022). Moreover, the fast developments in AI advancement often challenges the ability of teacher education programs to revise curricula, widen scope, and foster meaningful engagement with the technology (Hu et al., 2022). Multiple concerns were also raised with respect to the sustainability and broader applicability of AI-powered technologies in the context of the teaching-learning process within pre-service teacher institutions (Cáceres-Nakiche et al., 2024; Hu et al., 2022; Meylani, 2024).

Generally, there is a lack of systematic research on the application of AI in teacher education, as opposed to the arising trends of AI research in multiple settings and fields in education. The existing literature is varied in scope, methods, and focus areas, making it challenging to establish existing effects of AI on preservice teacher learning (Hu et al., 2022; Zawacki-Richter et al., 2019; Celik et al., 2022). Furthermore, there is a gap in inspecting the context and implementation factors relevant for enabling AI-supported decision-making in teacher education courses (Hu et al., 2022).

To address these existing gaps, this research conducts a meta-analysis to consolidate the existing evidence on AI use and its impact on preservice teacher learning outcomes. The meta-analysis identifies features of each study, including country of implementation, academic discipline, targeted learning outcomes, type of AI tools used, level of educational integration, duration, and scale of implementation. It also delineates the types of AI-based tools utilized and evaluates their impacts on the cognitive, affective, and psychomotor domains of preservice teachers. Additionally, the study examines the effectiveness of AI-based instruction within the consideration of different moderators. Drawing on evidence from multiple studies, this research provides practical strategies and informed guidance for the effective adoption of AI technologies in teacher education.

Literature Review

Artificial Intelligence

Artificial intelligence (AI) has become a significant transdisciplinary subject for industries and society and not just a complex topic in computer science. Early AI research was influenced by rule-based systems and symbolic reasoning, with the aim of simulating human problem-solving through explicit programming (Russell & Norvig, 2020). Nonetheless, the rise of machine learning, and specifically deep learning, has interchanged the framework to data-centric methodologies that allow machines to learn patterns and make decisions out of vast amounts of data without heavy explicit programming (LeCun, Bengio, & Hinton, 2015). This revolution has brought rapid advancement to fields such as computer vision, natural language processing, and robotics with useful applications ranging from driverless cars to virtual assistants.

One of the key drivers of recent AI breakthroughs is the creation of the model of architecture of the human brain to analyze intricate data hierarchically known as deep neural networks. Convolutional neural networks (CNNs) have transformed image recognition processes, while recurrent neural networks (RNNs) and transformers have improved natural language comprehension and generation (Vaswani et al., 2017). Introduction of large language models (LLMs) such as GPT-3 and GPT-4 has shown unprecedented ability to produce coherent and contextually appropriate text, opening doors to new opportunities in content generation, translation, and conversational AI (Brown et al., 2020). These models are trained with giant datasets and demand huge computational power and have raised questions regarding the ecological concern and availability of AI technologies.

Despite its positive impacts to society, utilization of AI encounters a variety of challenges and ethical issues. Algorithmic bias, in which AI systems were linked with societal biases existing in training datasets, poses a threat to fairness and equity (Barocas, Hardt, & Narayanan, 2019). Transparency and concreteness linger as major concerns, as some deep learning models are "black boxes," hindering users from understanding decision-making processes generated through AI (Doshi-Velez & Kim, 2017). In addition, privacy, security, and misuse risks of AI technologies have driven demands for stringent regulatory environments and responsible AI development processes (Floridi et al., 2018). These challenges need to be met to make sure that AI brings benefits to society without causing harm.

In the future, AI research will most likely evolve to construct more contextualized and human-centered intelligence. Current advancements in explainable AI, reinforcement learning, and multimodal models that combine vision, language, and reasoning aims to formulate systems that can exhibit more flexible and bounded behavior (LeCun, 2022). Moreover, interdisciplinary work is emerging, where AI develops interrelatedness with neurosciences, cognitive science, and ethics to establish a more reliable technology. As AI continuously evolves up to this day, its seamless inclusion in daily life continuously unfolds, which reveals its further strengths, weaknesses, and implication to the society.

Artificial Intelligence in Teacher Education

The use of artificial intelligence (AI) in teacher preparation has been a revolutionary force, transforming how teachers are trained and how they practice classroom teaching. AI-driven technologies such as Intelligent Tutoring Systems (ITS), analytics enabled by AI, and automated testing tools have brought in new opportunities for customized learning, professional development, and innovative pedagogical approaches. These tools combine the use of adaptive instruction to support individual teachers' requirements, develop critical thinking, and offer instant feedback that results in enhanced teaching-learning process (Celik et al., 2022). As AI continues to evolve, its impact on teachers' training enhances, and schools have been enabled to look back at conventional training methods and explore adopting technology-based approaches powered by AI tools (Alexandrowicz, 2024).

While there exists the possibility of AI, significant challenges exist in the implementation of AI in teacher training. The most significant challenge mentioned is teachers' demotivation in the implementation of AI, typically grounded in insufficient institutional support, insufficient guidance, and concerns over privacy and reliability (Celik et al., 2022). Studies indicate that less than a third of educators apply AI tools in actual practice, and for the most part, they blame this on a lack of practical experience and training as major barriers (Alexandrowicz, 2024). Furthermore, ethics such as algorithmic bias and data privacy remain critical issues that necessitate ongoing attention and the development of clearly established ethical principles for AI use in education (Zawacki-Richter et al., 2019).These issues demand in-depth, immersive, and contextualized training courses that address both the technical and ethics sides of AI deployment (Celik et al., 2022).

Best practices in teacher training to embed AI emphasize the need for tailored and cooperative professional development. Effective teacher training programs begin with principle-based fundamentals of AI, increasingly introducing technical details and examining the social and ethical implications of AI in education (Alexandrowicz, 2024). Providing teachers time for experiential trials using a range of AI tools—not just broadly accepted platforms such as ChatGPT—has also been shown to increase confidence and uptake of AI in teaching (Celik et al., 2022).In addition, effective programs address will, skill, and tool readiness, and knowledge and motivation, to ensure that teachers are prepared to work with and take advantage of AI technologies (Graduate Programs for Educators, 2023). The literature also refers to the importance of collaborative learning environments in which teachers can exchange experiences and approaches to AI utilization, enhancing professional development (Alexandrowicz, 2024).

Lastly, the literature calls for teacher education to focus on AI literacy and ethical awareness. With increasingly embedding of AI in the practice of teaching, teachers will need to be ready not just to utilize such technologies but also to critically examine their effect on teaching and learning (Celik et al., 2022). Quality teacher education programs with digital literacy, ethical consciousness, and real-world application are vital to supporting teachers to flourish within an AI-driven learning context (Zawacki-Richter et al., 2019). By solving the problems and embracing best practices, teacher training can tap the maximum potential of AI to create adaptive, effective, and fair teacher and student environments (Alexandrowicz, 2024).

Technology Integration (SAMR Model as Theoretical Framework)

To harmoniously incorporate the application of technology in the educational parlance, the educators may utilize the SAMR Model of Dr. Ruben Puentedura in the process of teaching-learning. It represents Substitution, Augmentation, Modification, and Redefinition. The first two tiers, the Substitution and Augmentation, emphasize the improvement of instruction, whereas the final two tiers, the Modification and Redefinition, emphasize the reformation of the learning experiences. This model encourages teachers to "move up" from lower to higher levels of learning involvement with technology. Whenever the integration results in a transforming experience, the higher levels of teaching and learning process can be felt by the teachers (Cáceres-Nakiche et al., 2024; Hamilton et al., 2016).

The initial stage of the model is Substitution, where an electronic and newer technology substitutes for an old technology. But there is no observed functional change in technology integration. For example, a teacher decides to use an electronic version of a mathematical test questionnaire, rather than a paper copy. Additionally, the second tier of the model is Augmentation where the technology is substituted with a more digital equivalent, but there are apparent functional enhancements in its use. The tool's functionality is enhanced in the support of the instructional task. For example, rather than manually graphing an equation on a graphing board, the students can think about entering the equation into a graphing calculator, which yields smoother and more precise results. At the Modification level, the third level of the model, the technology integration is centered on the task redesign relevant to the task. For instance, a science teacher can utilize an interactive computer simulation of light with variable set-ups that can be controlled rather than showing a printed diagram of light traveling. Use of scientific or mathematical digital software may be at this level. The most advanced level of the SAMR Model is the Redefinition level in which the integration of technology leverages contemporary tools that allow teachers and students to do things that they could not do previously. For instance, in conducting an experiment and recording the findings, the teacher requests the students to produce and present their observations and conclusions in the form of an edited video, rather than writing them on an experiment report (Cáceres-Nakiche et al., 2024; Hamilton et al., 2016).

Multiple studies have identified several benefits of applying the SAMR Model in the teaching-learning process. It provides a common framework for educators to reflect on and leverage their technology integration strategies, leading to a stronger professional growth. Fostering learner interest and participation, developing cooperation, and empowering differentiated instruction to enable teachers to construct personalized and tailor-fitted instruction for the diverse learners were the key features of the SAMR Model (Cáceres-Nakiche et al., 2024; Morales-Garcia et al., 2022; Tunjera & Chigona, 2020). In addition, the model's inclination towards the advancement from the enhancement levels (Substitution and Augmentation) to transformation levels (Modification and Redefinition) complements recent aims for education in the 21st century, including creativity, critical thinking, and digital literacy skills (Hamilton et al., 2016).

Despite its perceived benefits, the SAMR Model has come under scrutiny with respect to its hierarchical format and limitations in the technological application. Hamilton et al., (2016) posited that the model may not take pedagogical context or domain specificity very well into account. Furthermore, other studies have stated that teachers in general are contented to stay in the lower domains of the model, doing largely Substitution or Augmentation, and not necessarily moving towards transformational practice (Cáceres-Nakiche et al., 2024; Morales-Garcia et al., 2022). This requires ongoing professional development and continuous upskilling to appropriately apply emerging technologies (Hamilton et al., 2016).

SAMR Model is an excellent technology school integration theory. The structured, chronological progression of the model from Substitution through to Redefinition supports teachers to navigate their technology use toward more innovative, efficient teaching and learning processes. Purposeful use, however, relies on familiarity with the model's limitations as well as on reflective, situational practice (Hamilton et al., 2016; Cáceres-Nakiche et al., 2024).

Research Questions

The primary objective of this study is to evaluate the effectiveness of AI integration in enhancing preservice teachers' learning outcomes through a meta-analysis. Specifically, this study seeks to address the following questions:

1. How may the included studies be described in terms of the following:
    - 1.1. country of implementation;
    - 1.2. academic discipline;
    - 1.3. targeted learning outcomes;
    - 1.4. type of AI tool used;
    - 1.5. level of AI integration;
    - 1.6. duration of implementation, and;
    - 1.7. scale of implementation?
2. What types of AI-based interventions have been implemented to improve preservice teachers' learning outcomes ?
3. How effective are AI-based interventions in enhancing preservice teachers' learning outcomes across the following domains?
    - 3.1. Cognitive domain
    - 3.2. Affective domain

3.3. Psychomotor domain

4. How does the effectiveness of AI-based interventions in preservice teacher education vary, specifically in relation to:
   4.1. country of implementation;
   4.2. academic discipline;
   4.3. targeted learning outcomes;
   4.4. type of AI tool used;
   4.5. level of AI integration;
   4.6. duration of implementation, and;
   4.7. scale of implementation?

METHOD

Research Design

This study employed a meta-analytic research design, which quantitatively synthesizes the results of multiple empirical studies to determine the overall effect of artificial intelligence (AI)-based interventions on preservice teacher education. The choice of meta-analysis is methodologically appropriate for addressing the fragmented and heterogeneous nature of existing research on AI integration in teacher preparation programs. By aggregating effect sizes from various studies, this approach facilitates a broader understanding of AI's impact across cognitive, affective, and psychomotor learning domains. The design allows for comparisons between treatment (AI-based intervention) and control (traditional instruction) groups across different educational contexts.

This quantitative synthesis adheres to established standards for systematic reviews and meta-analyses (e.g., PRISMA guidelines), incorporating stringent procedures for literature search, study selection, coding, and statistical analysis. The methodological rigor increases the reliability and generalizability of the findings (Figure 1).

Literature Search Procedures

To ensure comprehensive coverage of relevant literature, an extensive search was conducted across several academic databases including OpenAlex, Semantic Scholar, Scopus, CrossRef, Google Scholar, and PubMed. Additionally, a manual search was performed to identify studies that may not have been indexed in the databases. The initial search yielded 5,880 records. After removing 1,143 duplicates electronically and 36 manually, 4,701 unique records were screened based on titles and abstracts. Of these, 2,503 were excluded for being unrelated, and another 2,160 were dismissed due to being non-empirical, developmental in nature, not written in English, or lacking methodological clarity. Full-text reviews were conducted on 38 articles, of which only 11 met all inclusion criteria, yielding a total of 19 extractable effect sizes for the final meta-analysis.

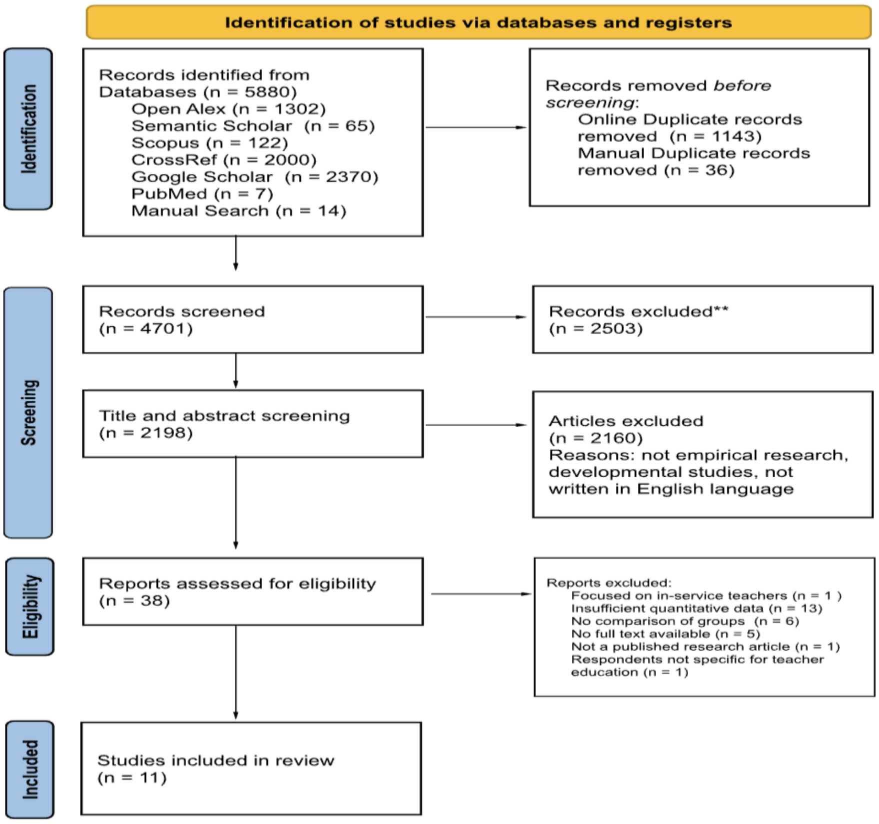

Figure 1
PRISMA flow diagram

Inclusion and Exclusion Criteria

Studies were included in the meta-analysis if they met the following criteria: (1) the study was an empirical investigation involving preservice teachers; (2) it implemented AI-based instructional or learning interventions; (3) it reported posttest quantitative data comparing experimental and control groups; (4) the publication was a full-length, peer-reviewed article or conference paper; and (5) the study was written in English. Studies were excluded if they focused solely on in-service teachers, did not include comparative groups, lacked extractable quantitative data, were inaccessible in full-text format, or were theoretical or developmental in nature. These criteria ensured that only rigorous, comparable, and relevant studies were included in the synthesis.

Coding Procedures

A structured coding protocol was used to extract both descriptive and quantitative information from each selected study. Descriptive variables included the author(s) and year of publication, country of implementation, academic discipline, targeted learning outcomes, learning domain (cognitive, affective, psychomotor), type of AI tool used, level of AI integration based on the SAMR model (Substitution, Augmentation), duration of intervention, and the scale of implementation (classroom-based, school-wide, or university-wide). Quantitative data were extracted in the form of posttest means, standard deviations, and sample sizes for both experimental and control groups. Multiple coders independently reviewed and verified the data to ensure reliability, with discrepancies resolved through discussion and consensus. This systematic process ensured a high level of consistency and validity in the dataset used for analysis.

Effect Size Calculation

To assess the impact of AI-based interventions, effect sizes were calculated using Hedges' g, which adjusts for small sample size bias and standardizes the mean difference between experimental and control groups. Only posttest data were used to isolate the effect of the intervention. Each study's effect size was accompanied by its standard error and 95% confidence interval. Due to significant heterogeneity among the included studies ($I^2 = 94.34\%$), a random-effects model was applied to estimate the overall effect size. Moderator analyses were conducted to explore the influence of country, learning outcome type, intervention duration, and level of technology integration on the observed effect sizes. Publication bias was assessed using a funnel plot and Egger's regression test, both of which indicated asymmetry and suggested a potential bias in favor of studies reporting significant results.

FINDINGS

Among the 5,880 papers initially retrieved from the literature search, a total of 11 studies qualified in the meta-analysis. Table 1 summarizes the included studies, displaying key information such as the authors and year of publication, country, academic discipline, targeted learning outcome, domain of learning, artificial intelligence (AI) used, level of technology integration, duration, scale of implementation, variable under consideration, and comparison between experimental and control group with pertinent statistical data.

General Study Characteristics

A total of 19 primary studies were included in this meta-analysis, covering both experimental and control groups across a variety of geographic, disciplinary, and pedagogical contexts under a rigorous inclusion and exclusion criterion. The studies primarily originated from Asia and the Middle East, reflecting a growing regional focus and integration in the role of artificial intelligence (AI) in pre-service teacher education. Specifically, China contributed the greatest number of studies (Li, 2023; Ji et al., 2023; Lu et al., 2024), with Malaysia and Ethiopia having two (Kumar, 2021; Elsayed et al., 2024) studies each. Moreover, a study was conducted in South Korea (Kim, 2024), Palestine (Younis, 2024), Jordan (Gasaymeh & AlMohtadi, 2024), Oman (Behforouz & Al Ghaithi, 2024), and the United Arab Emirates (Eltahir & Babiker, 2024), respectively. An individual study was also not identified with respect to its country of origin (Li & Ironsi, 2024). Identifying this diverse geographic distribution posits an emerging trend on the application of AI in pre-service teacher education across multiple countries and regions.

Table 1
Summary of studies in the meta-analysis

Author/s and Year of Publication	Country	Academic Discipline	Targeted learning outcome	Domain	AI tool used	Level of technology integration	Duration	Statistical Data					
								Conventional			Experimental		
								Posttest	Posttest				
								Mean	SD	Sample	Mean	SD	Sample
Kim (2024)	South Korea	history, math, informatics, biology, physics	Teaching Expertise	Psychomotor	*unreported*	Augmentation	15 weeks	3.50	0.40	13	3.82	0.50	26
Kumar (2021)	Malaysia	English	Learning Performance	Cognitive	Researcher-made Chatbot	Augmentation	10 weeks	39.933	2.572	30	42.50	2.675	30
	Malaysia	English	Perception of Learning	Affective		Augmentation	10 weeks	4.37	0.54	30	4.244	0.479	30
Younis (2024)	Palestine	computer science, math, languages, geography and history, science, religious	AI Literacy	Cognitive/ Psychomotor	ChatGPT	Augmentation	24 training-hours	2.297	0.4181	37	4.471	0.2152	37
Elsayed et al. (2024)	Ethiopia	language proficiency (English)	Demotivation	Affective	automated assessment platform	Augmentation	8 weeks	4.891	2.078	46	12.934	3.929	46
	Ethiopia	language proficiency (English)	Academic success	Cognitive	automated assessment platform	Augmentation	8 weeks	4.891	2.078	46	12.934	3.929	46
Lu (2024)	China	mathematics, science, computer	self-efficacy	Affective	ChatGPT	Substitution	4 weeks	3.73	0.46	107	3.97	0.43	107
	China	mathematics, science, computer	higher order thinking	Cognitive	ChatGPT	Substitution	4 weeks	3.37	0.27	108	3.8	0.27	108
Gasaymeh & AlMohtadi (2024)	Jordan	programming	programming skills (cognitive)	Cognitive	ChatGPT	Substitution	8 weeks	7.56	1.84	36	8.34	1.42	38
Behforouz & Al Ghaithi (2024)	Oman	language learning	Self-directed Learning	Affective	WhatsApp chatbot	Augmentation	4 weeks	1.153	12.802	25	2.313	23.285	25
Eltahir & Babiker (2024)	UAE	*Unreported*	academic performance	Cognitive	Moodle, Kahoot!, Studiosity	Substitution	4 months	15.31	1.53	55	18.32	1.21	55
Ji et al. (2023)	China	*Unreported*	STEM teaching literacy	Psychomotor	ChatGPT	Augmentation	8 weeks	81.27	1.554	15	81.29	1.589	21
	China	*Unreported*	Learning Performance	Cognitive	ChatGPT	Augmentation	8 weeks	68.67	3.055	15	85	4.243	21
Li & Ironsi (2024)	*unreported*	English	students' scores	Cognitive	ChatGPT	Substitution	14 weeks	62.28	6.019	29	74.05	4.072	29
Li (2023)	China	Modern Educational Technology	project performance	Cognitive	ChatGPT, Tencent QQ applications	Augmentation	3 weeks	87.27	3.41	39	88.8	2.73	42
	China	Modern Educational Technology	self-efficacy	Affective	ChatGPT, Tencent QQ applications	Augmentation	3 weeks	32.85	4.54	39	35.05	3.73	42
	China	Modern Educational Technology	learning attitudes	Affective	ChatGPT, Tencent QQ applications	Augmentation	3 weeks	30.00	4.21	39	31.5	2.76	42
	China	Modern Educational Technology	learning motivation	Affective	ChatGPT, Tencent QQ applications	Augmentation	3 weeks	24.98	2.09	39	26.17	2.73	42
	China	Modern Educational Technology	creative thinking tendency	Cognitive	ChatGPT, Tencent QQ applications	Augmentation	3 weeks	17.28	3.96	39	19.19	3.00	42

The included studies encompassed multiple academic disciplines, which exemplifies the flexibility on the use of AI across intertwined fields in the context of teacher education. Across these studies, multiple academic disciplines are covered in an individual paper. The most commonly sampled field was in language education (e.g., Kumar, 2021; Elsayed et al., 2024; Behforouz & Al Ghaithi, 2024; Li & Ironsi, 2024), computer programming and education (e.g., Gasaymeh & AlMohtadi, 2024; Younis, 2024; Lu et al., 2024), mathematics education (e.g., Kim, 2024; Lu et al., 2024), science education (e.g., Kim, 2024; Lu et al., 2024), and STEM education (Ji et al., 2023). Other areas included educational technology (e.g., Li, 2023; Ji et al., 2023), and social science education (e.g., Kim, 2024), which leads to a more specific application of AI in mentoring pre-service teachers. Some studies (e.g., Ji et al., 2023; Li, 2023) were interdisciplinary in nature, integrating AI tools with a wider focus on developing pre-service teachers' teaching strategies and other aptitudes leading to the development of their pedagogy. The diverse nature of the inclusion of the AI in the pre-service teacher education showcases the instruction which equips pre-service teachers with appropriate 21st-century skills and encouraging contextual technology integration and across subject areas.

Within the context of the learning domains, the covered studies included outcomes in the cognitive (e.g., Kumar, 2021; Younis, 2024; Li, 2023), affective (e.g., Lu et al., 2024; Behforouz & Al Ghaithi, 2024), and psychomotor (e.g., Ji et al., 2023; Kim, 2024) domains. Outcomes measured cognitively such as academic performance, learning achievement, and higher-order thinking were most commonly assessed, while affective outcomes covered constructs like self-efficacy, motivation, and perception. Some studies covered psychomotor skills, principally those linked to teaching practice and technical competencies for teacher education such as teaching literacy.

In addition, the included studies utilized different AI tools to strengthen instruction for pre-service teachers. Among the AI tools, ChatGPT was the most common (e.g., Ji et al., 2023; Lu et al., 2024; Li, 2023), signifying its extensive application in educational contexts. Other studies harnessed researcher-designed chatbots (Kumar, 2021), WhatsApp-integrated AI systems (Behforouz & Al Ghaithi, 2024), automated assessment tools (Elsayed et al., 2024), Kahoot!, Moodle, Tencent QQ, and Studiosity (Li, 2023; Eltahir & Babiker, 2024), among others. These tools were embedded in relation to the multiple levels of the SAMR model, with the majority used within the augmentation level (e.g., Kim, 2024; Elsayed et al., 2024), where AI enhanced traditional learning activities with an identified functional change. Fewer studies used AI at the substitution level (e.g., Lu et al., 2024; Li & Ironsi, 2024), merely replacing traditional methods without functional change. No study is identified which applied AI within the modification or redefinition levels.

In terms of the implementation duration of AI tools, a variation across studies is observed, with a range of less than 3 weeks up to 15 weeks, with the majority falling within 4 to 10 weeks (e.g., Lu et al., 2024; Kumar, 2021; Ji et al., 2023). The scope of the utilization was also varied, where majority of the studies applied AI at the university level (e.g., Younis, 2024; Li & Ironsi, 2024), a few at the classroom level (e.g., Li, 2023), while others encompass entire schools (e.g., Behforouz & Al Ghaithi, 2024;

Eltahir & Babiker, 2024), and some did not report the specified scale (e.g., Ji et al., 2023). Applying AI tools in multiple settings embodied its flexibility and variability within the teacher education parlance.

All 19 studies applied experimental setups with control and experimental groups (e.g., Kim, 2024; Elsayed et al., 2024; Li, 2023), as explicitly mentioned in the inclusion criteria. Across the studies, AI-led interventions generally led to higher posttest means in the experimental groups compared to the control groups, confirming the positive impact of AI integration on targeted educational outcomes in teacher education. Moreover, the subsequent findings are based on identified studies through a rigorous and well-structured inclusion and exclusion criteria, which ensured the robustness of data used in the meta-analysis. In addition, the redefining potential of AI in teacher education was exemplified in these studies. These studies justify the capability of AI tools to enrich learning outcomes across three learning domains, enhance professional competencies, and facilitate modern pedagogies among future educators. These insights highlight the importance of empirical approaches in integrating AI into teacher education courses.

Table 2
AI tools and their applications in pre-service education settings

AI Tool Used	Number of Effect Sizes	Instructional Strategies Used	Study Exemplar
ChatGPT	12	Modular Learning, Hands-on Learning, Task Design, Flipped Learning, Collaborative Learning,	Lu (2024) employed the ChatGPT-supported training approach where preservice teachers used ChatGPT alongside these tools for lesson planning. In particular, the researchers created four tasks using ChatGPT as the learning material for the experimental group. The four tasks were mainly on instructional design, with simulated classroom activities as additional exercises. Instructional design encompasses planning, designing, hypothesizing, and organizing classroom instruction, and it is a necessary part of lesson preparation for instructors. ChatGPT can help preservice teachers in generating and fine-tuning instructional designs according to instructions and hence enhance their lesson preparation effectiveness and teaching performance
Tencent QQ	5	Flipped Learning	Li (2023) utilized a flipped instruction approach where the instructor delivered knowledge on courseware designing and developing to all the students, including Microsoft PowerPoint operations and theories.. In particular, the experimental group followed ChatGPT-based flipped learning guiding approach (ChatGPT-FLGA) for 3 weeks.
Automated Assessment Platform	2	Feedbacking, Active Learning, Metacognitive Learning, Mentoring	Elsayed, et al. (2024) allowed educators to monitor students' progress actively through the dashboard of the AI platform and offer real-time guidance by answering students' queries, clarifying misconceptions, and providing motivational assistance. Educators made sure the students knew how to use the AI tools and the feedback offered by the system to their advantage. Besides exam support, teachers in the experimental group conducted regular review sessions. Teachers were also available for individual consultations with the students.
WhatsApp Chatbot	1	Self-directed learning	Through Python language programming, an interactive chatbot was developed by Behforouz and Al Ghaithi (2024) using WhatsApp to be used in this study among the experimental group participants. The experimental group was given all the instructions, explanations, and tests via the chatbot. The researchers input the necessary tasks and the instructions in the chatbot database and update them from time to time according to the activities students need to do.
Moodle, Kahoot!, Studiosity	1	Gamification. Distance Learning	Eltahir and Babiker (2024) explored that when carefully integrated into the educational system, advanced AI-powered products can multiply the effectiveness of the educational environment. Among the many means, one approach to integrating Chatbots, especially OpenAI's GPT-3 for Learning Into moodle LMS, during instruction for teaching the learning of design topics. These Chatbots assisted learners in creating study plans and informative materials based on ADDIE and ASSURE

			instructional design frameworks. Additionally, it includes many gamified learning platforms, including Kahoot! These AI-based gamification tools reshape traditional learning processes by integrating fun elements. Kahoot! was used to create interactive tests, discussions, and polls aimed at creating a fun and competitive atmosphere for students. Additionally, some online assistance and support facilities are available using tools like Studiosity.
Researcher-made	1	Collaborative Learning, Metacognitive Learning	The QMT212 chatbot, christened after the course code, was created by Kumar (2021) as a teaching assistant for an instructional design course to facilitate group-based, project-based learning within a ten-week period. Students worked in groups of five, and ten Educational Chatbots (ECs) were used to facilitate teamwork and organize course activities. EC1 to EC4—Welcome Bot, Group Registration Bot, Group Leader Registration Bot, and Project Registration Bot—processed administrative functions. EC5 (Assignment Bot) and EC6 (Picaso Bot) aided assignment processes, EC7 (Perception Bot) collected feedback from users and measured acceptance of the chatbot, EC8 (Progress Bot) monitored teamwork progress, EC9 (Report Writing Bot) served as a project guide and FAQ, and EC10 (Peer-to-Peer Evaluation Bot) aided peer evaluation. Each of the chatbots was programmed based on micro-learning principles in order to avoid cognitive overload and ensure effective interaction, all aimed at encouraging collaborative work on a project that entailed designing an e-learning tool, producing a detailed report, and giving a final presentation.
Unreported	1	Exploratory Learning, Lesson Study/ Microteaching	Kim (2024) utilized a step-by-step analyze-explore-design-implement-evaluate process. In the analysis phase, participants had identified teaching problems, with the aim of incorporating AI to solve classroom problems. In the exploration phase, they studied TPACK theory, curriculum samples, and AI tools for lesson design. In the design phase, they organized identified issues and developed AI-integrated lesson plans from their exploration. The application phase was in the form of microteaching sessions where participants rehearsed presenting their AI-augmented lessons. Lastly, during the evaluation stage, students compared and evaluated lessons, noting areas of improvement.

Table 2 presents an overview of different AI tools incorporated into teaching strategies in several studies. ChatGPT was the most frequently used, facilitating modular, hands-on, flipped, and collaborative learning through instructional design tasks. Tencent QQ and a flipped learning guide based on ChatGPT were also used to instruct courseware development. An automated testing platform helped to monitor progress by students and provide individualized feedback. A WhatsApp chatbot facilitated autonomous learning through Python-coded interactions. Moodle, Kahoot!, and Studiosity were some other tools which facilitated gamified and remote learning. Moreover, the researcher-developed QMT212 chatbot served as an instructional aid for a group-focused instructional design course, using ten expert bots borrowed from micro-learning to help with administration, assignment, progress monitoring, and peer assessment. One study focused on exploratory learning using a lesson study approach of incorporating AI for planning instructions and microteaching.

Table 3
Overall effect size and heterogeneity analysis

	k	ES (g)	SE	Variance	95% CI		z	p	Q	df(Q)	p	I^2
					Lower	Upper						
Fixed	19	0.978	0.055	0.003	0.870	1.086	17.752	0.000	318.124	18	0.000	94.342
Random	19	1.314	0.238	0.057	0.848	1.780	5.523	0.000				

Note. k=number of effect sizes; g=Hedges' g; SE=standard error; CI=confidence of interval for the average value of ES; Q=Homogeneity Value; df=degrees of freedom; I^2=level of heterogeneity

Table 3 displays the average effect size, heterogeneity value, and confidence intervals according to the effect model in the analysis. It can be gleaned from Table 2 that the heterogeneity analysis was significant ($p < .05$). The Q value was identified as 318.124

with degrees of freedom of 18, implying that the studies included in the meta-analysis are significantly heterogeneous and do not share a similar effect size (Borenstein et al., 2009). This leads to the utilization of the random-effect size model. Moreover, the level of heterogeneity displayed a value of 94.342%, which led to a high heterogeneity across 19 effect sizes. Consequently, the moderator analysis is relevant to be conducted (Higgins & Thompson, 2002). The computed effect sizes ranged from 0.848 (lower limit) to 1.780 (upper limit) from the random-effects model within a 95% confidence interval. The overall weighted effect size of 1.314 substantiates that the use of artificial intelligence leads to a significantly large and positive effect (Cohen, 1988) on pre-service teachers' learning outcomes, as evidenced in multiple studies (e.g., Younis, 2024; Elsayed et al., 2024; Li & Ironsi, 2024).

	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-value
Behforouz & Al Ghaithi (2024)	0.061	0.278	0.078	-0.485	0.607	0.218	0.827
Elsayed et al. (2024a)	2.538	0.279	0.078	1.991	3.084	9.101	0.000
Elsayed et al. (2024b)	2.538	0.279	0.078	1.991	3.084	9.101	0.000
Eltahir & Babiker (2024)	2.167	0.239	0.057	1.698	2.636	9.061	0.000
Gasaymeh & Al Mohtadi (2024)	0.471	0.233	0.054	0.014	0.929	2.019	0.043
Ji et al. (2023a)	0.012	0.331	0.109	-0.635	0.660	0.038	0.970
Ji et al. (2023b)	4.203	0.595	0.355	3.036	5.370	7.058	0.000
Kim (2024)	0.667	0.341	0.116	-0.002	1.336	1.955	0.051
Kumar (2021a)	0.966	0.270	0.073	0.437	1.494	3.581	0.000
Kumar (2021b)	-0.244	0.256	0.065	-0.745	0.258	-0.952	0.340
Li & Ironsi (2024)	2.260	0.333	0.111	1.606	2.913	6.778	0.000
Li (2023a)	0.493	0.224	0.050	0.054	0.981	2.203	0.028
Li (2023b)	0.526	0.224	0.050	0.087	0.966	2.349	0.019
Li (2023c)	0.421	0.223	0.050	0.016	0.857	1.888	0.059
Li (2023d)	0.482	0.223	0.050	0.044	0.920	2.159	0.031
Li (2023e)	0.541	0.224	0.050	0.102	0.981	2.413	0.016
Lu (2024a)	0.537	0.139	0.019	0.265	0.809	3.873	0.000
Lu (2024b)	1.587	0.156	0.024	1.282	1.892	10.198	0.000
Younis (2024)	6.470	0.579	0.336	5.334	7.606	11.166	0.000
Fixed Pooled	0.978	0.055	0.003	0.870	1.086	17.752	0.000

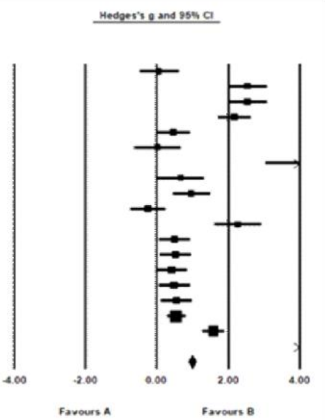

Figure 2
Forest plot showing the distribution of effect sizes of the studies (n=19)

Figure 2, or the forest plot distribution of Hedges' g effect sizes, shows that all of the studies included in the meta-analysis leaned towards the experimental groups exposed to artificial intelligence (AI) tools over the control group that received traditional instruction. Upon inspecting each of the studies, the maximum effect size was g = 6.470 (Younis, 2024), while the minimum effect size was g = –0.244 (Kumar, 2021). Fifteen (15) studies displayed a statistically significant p-value ($p < .05$), where there are significant differences between the experimental and control groups with respect to their posttest mean scores across multiple learning outcomes exhibited by pre-service teachers (i.e., Elsayed et al., 2024; Eltahir & Babiker, 2024; Ji et al., 2023; Lu et al., 2024; Li, 2023).

Moderator analysis of studies

The use of moderator analysis in this meta-analysis aims to determine whether study features might be responsible for the effect size variation of AI interventions in pre-service teacher education. Therefore, the between-group Q statistic (*Qb*), a homogeneity test, was used to ascertain whether there was a significant variation among subgroups for each moderator. A *Qb* value that is statistically significant ($p < .05$) indicates that the moderator effectively explains the heterogeneity in effect sizes across the studies.

Table 4
Moderator analyses

Moderator	k	Effect size (g)	95% CI LL	95% CI UL	Q_b	p
Country					177.633	0.000
China	9	0.847	0.398	1.297		
Ethiopia	2	2.559	2.169	2.949		
Jordan	1	0.476	0.014	0.939		
Malaysia	2	0.353	-0.838	1.563		
Oman	1	0.062	-0.493	0.616		
Palestine	1	6.538	5.391	7.686		
South Korea	1	0.681	-0.002	1.364		
UAE	1	2.182	1.710	2.654		
unreported	1	2.291	1.628	2.953		
Learning Outcome					130.374	0.000
AI Literacy	1	6.538	5.391	7.686		
Information Literacy	2	1.082	0.058	2.107		
Learning Achievement	6	2.052	1.173	2.931		
Learning Attitudes	1	0.425	-0.016	0.865		
Learning Motivation	2	1.516	-0.514	3.547		
Learning Perception	1	-0.247	-0.755	0.261		
Pedagogical Competence	2	0.342	-0.313	0.996		
Self-Directed Learning	1	0.062	-0.493	0.616		
Self-Efficacy	2	0.537	0.305	0.769		
Technology Literacy	1	0.476	0.014	0.939		
Level of Technology Integration					0.019	0.891
Augmentation	14	1.327	0.693	1.960		
Substitution	5	1.393	0.678	2.109		
Duration of implementation					94.849	0.000
Less than 3 weeks	1	6.538	0.253	0.956		
3-6 weeks	8	0.605	5.391	7.686		
More than 6 weeks	10	1.535	0.777	2.293		

The first moderator, country, yielded a statistically significant *Qb* value (*Qb* = 177.633, $p = .000$), indicating that the effect of AI interventions significantly differed depending on the country where the study was conducted. This heterogeneity emphasizes how national context influences the way AI is incorporated into teacher preparation programs. The effect sizes ranged widely, with Palestine exhibiting the highest impact ($g = 6.538$; Younis, 2024), followed by Ethiopia ($g = 2.559$; Elsayed et al., 2024), and the United Arab Emirates ($g = 2.182$; Eltahir & Babiker, 2024). Conversely, minimal effects were observed in Oman ($g = 0.062$; Behforouz & Al Ghaithi, 2024) and Malaysia ($g = 0.353$; Kumar, 2021), with confidence intervals that include zero, implying statistically non-significant results.

The learning outcome as the second moderator also yielded a significant *Qb* value (*Qb* = 130.374, $p = .000$), indicating that the kind of learning outcome the intervention was intended to achieve plays a significant role in the observed variation in effect sizes. AI literacy had the largest effect (*g* = 6.538; Younis, 2024), followed by learning achievement (*g* = 2.052; Elsayed et al., 2024; Li, 2023) and information literacy (*g* = 1.082; Ji et al., 2023). Learning motivation was moderately effective (*g* = 1.516; Li, 2023), although with some uncertainty. However, outcomes like self-directed learning (*g* = 0.062; Behforouz & Al Ghaithi, 2024), pedagogical competence (*g* = 0.342; Ji et al., 2023), and learning perception (*g* = –0.247; Kumar, 2021) were linked to small or negative effects, suggesting that AI's influence on affective and metacognitive domains is more constrained.

On the other hand, there was no discernible difference in effect sizes between the moderator level of technology integration, which was classified using the SAMR model (Substitution and Augmentation). This level did not produce a significant *Qb* value (*Qb* = 0.019, *p* = .891). Substitution studies (e.g., Lu et al., 2024; Li & Ironsi, 2024) reported a mean effect size of 1.393, while augmentation studies (e.g., Kumar, 2021; Ji et al., 2023; Kim, 2024) reported a mean of 1.327, both indicating positive effects. This suggests that pedagogical quality and instructional design may be more influential than SAMR levels per se.

Lastly, the duration of implementation of AI interventions was found to be a major moderator (*Qb* = 94.849, *p* = .000), confirming that this accounts for notable variance in results. Interventions lasting less than three weeks had the largest effect size (*g* = 6.538; Younis, 2024), likely due to intensive short-term designs or novelty effects. Interventions over six weeks also showed significant effects (*g* = 1.535; Elsayed et al., 2024; Kim, 2024), while those between three to six weeks had smaller effects (*g* = 0.605; Lu et al., 2024), suggesting possible transitional adaptation phases.

Publication Bias

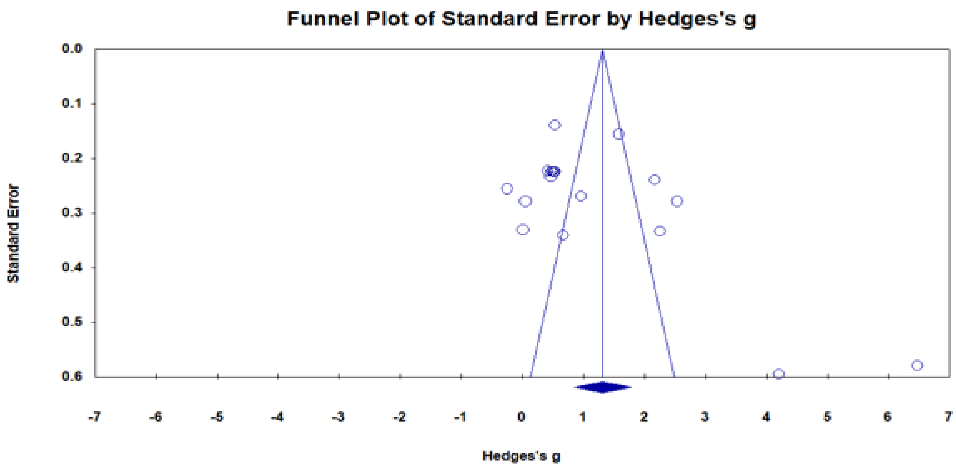

Figure 3
Funnel plot of standard error by hedges’s

Table 5
Egger's regression intercept

Intercept	5.82631
Standard error	2.87297
95% lower limit (2-tailed)	-0.23492
95% upper limit (2-tailed)	11.88754
t-value	2.02804
df	17.00000
p-value (1-tailed)	0.02926
p-value (2-tailed)	0.05852

Displayed by the funnel plot is the distribution of individual study effect sizes against their corresponding standard errors. If there is no publication bias and all studies are measuring the same true effect, the points on the plot should form a symmetrical, inverted funnel around the overall average effect. However, the funnel plot in this meta-analysis is noticeably asymmetrical, with most studies clustered on the right-hand side. This suggests that AI-based interventions have moderate to large positive effects, especially as reported by studies such as Younis (2024), Elsayed et al. (2024), and Li & Ironsi (2024).

Conversely, the studies with smaller or even negative effects are expected on the sparsely populated left-hand side of the plot. For instance, Kumar (2021) reported a small or even negative effect size ($g = –0.244$) for affective learning outcomes, which contributed to the sparse distribution. This asymmetry indicates the underrepresentation of published literature that yields null or non-significant findings, and hence, suggests potential publication bias. It might also reflect selective reporting trends, where studies with statistically significant or favorable outcomes are more likely to be published.

To statistically verify the observed asymmetry, Egger's regression test was conducted. As shown in Table 5, the intercept value was 5.826 with a standard error of 2.873, and the 95% confidence interval ranged from –0.235 to 11.888. Although the interval slightly overlaps zero, indicating some uncertainty, the test still showed a t-value of 2.028 ($df = 17$), with a one-tailed p-value of 0.029 and a two-tailed p-value of 0.059. The one-tailed result is statistically significant at the 0.05 level, indicating evidence of funnel plot asymmetry consistent with publication bias (Egger et al., 1997). The two-tailed result, while marginal, still reflects a visible trend toward asymmetry.

Moreover, the relatively high intercept value strengthens the concern. In the absence of publication bias, the intercept would be expected to be closer to zero. These results imply that smaller studies (i.e., Behforouz & Al Ghaithi, 2024; Ji et al., 2023) may be disproportionately associated with larger and more favorable effects, an outcome often attributed to selective publication practices in the literature on AI-enhanced teacher education.

To minimize bias in study selection, it is important to highlight that this meta-analysis utilized a rigorous and clearly defined set of criteria, appraising multiple databases to include both published and gray literature. Furthermore, each study (i.e., Li, 2023; Kim, 2024; Eltahir & Babiker, 2024) was evaluated for methodological quality, relevance to

AI interventions in pre-service teacher education, and the availability of extractable quantitative data.

Despite the careful procedures, the underrepresentation of small-scale or null-effect studies may reflect broader patterns within the educational research landscape. Thus, the statistical signals of potential publication bias—as revealed through Egger's test and visualized in the funnel plot—cannot be entirely dismissed. Meaning, while the meta-analysis was methodologically sound and transparent in its selection criteria, the existing body of literature may inherently favor significant results.

DISCUSSION

This meta-analysis provides strong evidence for including Artificial Intelligence (AI) in preservice teachers' training, with a large overall effect size ($g = 1.314$) attesting to the extreme positive impact of AI-enhanced teaching on learning outcomes. Applications such as ChatGPT, computerized assessment tools, and blended learning platforms have been found effective in promoting cognitive, affective, and psychomotor development among pre-service teachers (Younis, 2024; Elsayed et al., 2024; Li & Ironsi, 2024). The technologies appear to enhance instruction through instant feedback, customized learning, and greater content learning. Yet, the strong level of heterogeneity observed across the studies ($I^2 = 94.34\%$) emphasizes the need to examine contextual variables and implementation factors that are responsible for such differences in effectiveness.

The moderator analyses yielded substantial differences based on the country of implementation. Studies in Palestine (Younis, 2024), Ethiopia (Elsayed et al., 2024), and the United Arab Emirates (Eltahir & Babiker, 2024) exhibited the largest effect sizes, which would signify that positive national policy, infrastructure enablement, and institutional openness might have facilitated more easily the introduction of AI successfully into teacher training. On the contrary, limited reported effects in Malaysia (Kumar, 2021) and Oman (Behforouz & Al Ghaithi, 2024) might be indicative of context-related limitations such as limited technical capacities, less advanced digital pedagogical practice, or lacking faculty training. These findings strengthen the argument that while AI has global applicability, its impact relies significantly on the specific socio-educational context within which it is used.

The category of learning outcomes addressed also emerged as a major moderator, with interventions aimed at promoting AI literacy (Younis, 2024), learning attainment (Elsayed et al., 2024), and higher-order or creative thinking skills (Li, 2023; Lu et al., 2024) having the greatest effect sizes. These learning outcomes are naturally aligned with the strengths of AI in delivering structured content, processing student responses, and enabling adaptive learning. By contrast, less effective or insignificant impacts were associated with interventions on self-directed learning (Behforouz & Al Ghaithi, 2024), pedagogical competency (Ji et al., 2023), and learning perception (Kumar, 2021), which rely more on reflective thought, interpersonal communication, and situated practice—areas in which current AI solutions have limited capacity.

The study also probed to what degree AI integration was in terms of the SAMR model, distinguishing between substitution and augmentation. The results indicate no statistically significant difference in effect size between these two levels (Lu et al., 2024; Kim, 2024), suggesting that the degree of technological augmentation, as this has been theorized in SAMR, may not be the most important determinant of instruction effect. Conversely, pedagogical design quality, suitability of AI usage, and instructional alignment may be more crucial determinants. Moreover, insufficient research at the redefinition and modification levels indicates that groundbreaking applications of AI in teacher education have not been adequately analyzed in the empirical literature.

Implementation duration was also an important factor. Interventions that lasted less than three weeks produced the largest effect sizes, which may be due to heightened motivation or novelty effects (Younis, 2024). Interventions of more than six weeks also yielded strong positive results (Kim, 2024; Elsayed et al., 2024), which emphasizes the importance of ongoing interaction with AI technology for long-term learning gains. Three- to six-week interventions, however, had quite humble effects reported by them (Lu et al., 2024; Ji et al., 2023), maybe because they represent the period of acclimatization when teachers and students were becoming accustomed to AI technologies' implementation.

Finally, publication bias, as demonstrated by the asymmetrical funnel plot and Egger's test for regression, needs to be kept in mind. The data show that the literature may be underestimating small or null effects studies (Kumar, 2021; Ji et al., 2023) and, as a result, exaggerating the effectiveness of AI-based interventions as reported. Despite applying stringent inclusion criteria to this meta-analysis, the tendency of the literature to publish predominantly positive outcomes remains a matter of concern. The future of further research needs to be more open reporting, such as the publication of non-significant or mixed results, to establish a more balanced and reliable evidence base.

Overall, the integration of AI in preservice teacher education is of great promise, but it relies on a variety of factors such as national context, learning objectives, and implementation design. Cognitive and performance outcomes are closely related to AI affordances, but further exploration is needed to ascertain how AI can enable more advanced pedagogical and reflective capacities. Further, there ought to be exploration of higher-order levels of technology use and methodological openness to inform a more advanced and fuller theory of AI in transforming teacher education.

CONCLUSION

This meta-analysis combined data from 19 studies of the effectiveness of Artificial Intelligence (AI)-based interventions in enhancing preservice teachers' learning outcomes. The results reveal a large overall effect size ($g = 1.314$) in favor of AI-based interventions, which indicates that AI has a strongly positive influence on the cognitive, affective, and psychomotor dimensions of learning among preservice teachers. The data also confirms that not only are AI tools more than comparable to the conventional instruction methods, but they are also flexible and context-sensitive applications in different education settings and fields.

AI is generally being applied at the augmentation level of the SAMR model, and no research has yet reached the modification or redefinition levels. This means that AI is being used to enhance existing teaching-learning interaction, as opposed to something that transforms pedagogical planning and curriculum. The most used AI tool in the 19 studies is the ChatGPT, which indicates how simple it is to use and deploy in various pre-service teacher education contexts. Cognitive outcomes were treated most and evidenced high positive effects, followed by affective and psychomotor domains. The impact of AI on affective outcomes like motivation and perception was more mixed, however. Geographic and contextual influences strongly moderated the efficacy of AI interventions. For instance, those conducted in Palestine, Ethiopia, and the UAE yielded higher effect sizes, possibly because of national agendas or institutional preparedness for AI implementation. Duration of implementation matters, with short-term (<3 weeks) and long-term (>6 weeks) interventions showing larger effect sizes than mid-range durations (3–6 weeks), suggesting novelty and ongoing exposure. Although the promising findings, evidence of publication bias was indicated, with a trend towards reporting positive results in the specialty.

Overall, the research confirms that AI is highly capable of shaping the destiny of teacher training, especially if custom-made to desired learning results and applied in context and pedagogy sensitive manner. Nevertheless, the subject continues to unfold, and broader adoption—particularly at transformative SAMR levels—is less explored.

RECOMMENDATIONS

This meta-analysis identified that pre-service teacher education programs need to go beyond the Augmentation level and explore the possibility of enabling the Modification and Redefinition level with respect to the SAMR Model. The elements of the curriculum must foster interdisciplinary application and embed AI literacy into the professional or specialization courses to equip preservice teachers with the necessary competencies as major consumers and producers of knowledge, including a major catalyst as feedback enablers of AI tools. It is recommended that policymakers and education officials promote long-term AI adoption through continued investments in infrastructure, training, and equitable access, especially in low-resourced areas. Moreover, clear pedagogical and ethical guidelines should be developed to shape responsible AI utilization in teacher preparation. Future studies need to target understudied learning gains like pedagogical competence and self-directed learning, examine long-term and change-oriented AI interventions, and rectify publication bias by promoting dissemination of both noteworthy and null results to provide a balanced evidence base.

LIMITATIONS

While this meta-analysis provides informative details about the role of Artificial Intelligence in shaping the future of teacher education, various limitations need to be highlighted. Firstly, the study was conducted based on a relatively limited number of primary studies meeting the eligibility criteria (n = 19), which may limit the generalizability of the findings. Although an initial search of the literature proved to be well over 1,000 studies, the final group was constrained by what quantitative data were extractable, what intervention clarity had been reported, and whether control and experimental group comparison existed. This suggests that the broader research landscape for AI in teacher education is maybe not yet fully developed, with many studies being exploratory or less than adequately methodological to warrant inclusion in meta-analytic synthesis. Second, included studies were most concentrated in the Middle East and Asia, and China, Malaysia, and Ethiopia were overrepresented, whereas Western or underrepresented areas such as Africa, Latin America, and parts of Europe were underrepresented or missing. Therefore, the findings may capture regional trends and contextual factors rather than global variations in AI integration, infrastructure, or teacher education models. Third, while the research investigated AI integration from the perspective of the SAMR model, most included studies were at the levels of substitution and augmentation. No study looked at AI implementation at modification and redefinition levels, which hampered the study in evaluating the actual transformative capacity of AI in teacher education. In addition, a lack of routine reporting in some studies resulted in them not clearly defining the scale, length, or instructional practices involved in their AI interventions, making it impossible to perform more detailed contextual comparisons. Fourth, extreme heterogeneity (I^2 = 94.34%) was found between studies, and although the moderator analysis facilitated the identification of factors such as country, learning outcome, and length of intervention, the existence of such heterogeneity indicates that other untested variables (e.g., readiness of institutions, teacher digital literacy, or the quality of AI tools) potentially shape effectiveness as well. Third, the funnel plot shows asymmetry and is supplemented by Egger's regression result as well. Negative or null studies may get underreported within the literature so that the result would be positively skewed towards a more favorable explanation for the efficiency of AI. Despite these limitations, this study contributes significantly to the growing empirical evidence base for the incorporation of AI in teacher education and highlights key areas that need more empirical research.

REFERENCES

Included studies in the review

Alexandrowicz, M. (2024). Artificial intelligence in teacher education: Challenges and opportunities. *Journal of Educational Technology & Society, 27*(1), 45-60.

Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and machine learning. fairmlbook.org. https://fairmlbook.org/

*Behforouz, B., & Al Ghaithi, A. (2024). The Impact of Using Interactive Chatbots on Self-Directed Learning. *Studies in Self-Access Learning Journal*, *15*(3).

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165. https://arxiv.org/abs/2005.14165

Cáceres-Nakiche, K., Carcausto-Calla, W., Yabar Arrieta, S. R., & Lino Tupiño, R. M. (2024). The SAMR Model in Education Classrooms: Effects on Teaching Practice, Facilities, and Challenges. *Journal of Higher Education Theory and Practice, 24*(2), 160-172.

Cáceres-Nakiche, K., Carcausto-Calla, W., Yabar Arrieta, S. R., & Lino Tupiño, R. M. (2024). The SAMR Model in Education Classrooms: Effects on Teaching Practice, Facilities, and Challenges. *Journal of Higher Education Theory and Practice, 24*(2), 160-172. https://articlegateway.com/index.php/JHETP/article/view/6816

Celik, B., Yilmaz, R., & Yilmaz, M. (2022). The role of artificial intelligence in teacher professional development: A systematic review. *Computers & Education, 181*, 104438. https://doi.org/10.1016/j.compedu.2022.104438

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The Promises and Challenges of Artificial Intelligence for Teachers: a Systematic Review of Research. *TechTrends, 66*, 1069–1082. https://doi.org/10.1007/s11528-022-00715-y

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. https://arxiv.org/abs/1702.08608

*Elsayed, A. M., Kholikov, A., Abdullayeva, I., Al-Farouni, M., & Wodajo, M. R. (2024). Teacher support in AI-assisted exams: an experimental study to inspect the effects on demotivation, anxiety management in exams, L2 learning experience, and academic success. *Language Testing in Asia*, *14*(1), 53.

*Eltahir, M. E., & Babiker, F. M. E. (2024). The influence of artificial intelligence tools on student performance in e-learning environments: Case study. *Electronic Journal of e-Learning*, *22*(9), 91-110.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

*Gasaymeh, A. M. M., & AlMohtadi, R. M. (2024). The Effect of Flipped Interactive Learning (FIL) Based on ChatGPT on Students' Skills in a Large Programming Class. *International Journal of Information and Education Technology*, *14*(11).

Graduate Programs for Educators. (2023). Preparing teachers for AI integration: Best practices and strategies. *International Journal of Teacher Education, 15*(2), 112-130.

Hamilton, E. R., Rosenberg, J. M., & Akcaoglu, M. (2016). The Substitution Augmentation Modification Redefinition (SAMR) Model: A Critical Review and Suggestions for Its Use. *Tech Trends, 60*(5), 433–441. https://www.learntechlib.org/p/175669/

Hu, X., Salas-Pilco, S. Z., & Xiao, K. (2022). Artificial Intelligence and Learning Analytics in Teacher Education: A Systematic Review. *Education Sciences, 12*(8), 569. https://doi.org/10.3390/educsci12080569

*Ji, Y., Zou, X., Li, T., & Zhan, Z. (2023, November). The effectiveness of ChatGPT on pre-service teachers' STEM teaching literacy, learning performance, and cognitive load in a teacher training course. In *Proceedings of the 2023 6th international conference on educational technology management* (pp. 16-22).

*Kim, S. W. (2024). Development of a TPACK Educational Program to Enhance Pre-service Teachers' Teaching Expertise in Artificial Intelligence Convergence Education. *International journal on advanced science, engineering & information technology*, *14*(1).

*Kumar, J. A. (2021). Educational chatbots for project-based learning: investigating learning outcomes for a team-based design course. *International journal of educational technology in higher education*, *18*(1), 65.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

*Li, H. (2023). Effects of a ChatGPT-based flipped learning guiding approach on learners' courseware project performances and perceptions. *Australasian Journal of Educational Technology*, *39*(5), 40-58.

*Li, Z., & Ironsi, C. S. (2024). The efficacy of generative artificial intelligence in developing science education preservice teachers' writing skills: An experimental approach. *Journal of Science Education and Technology*, 1-12.

*Lu, J., Zheng, R., Gong, Z., & Xu, H. (2024). Supporting teachers' professional development with generative AI: The effects on higher order thinking and self-efficacy. *IEEE transactions on learning technologies*.

Meylani, R. (2024). Artificial Intelligence in the Education of Teachers: A Qualitative Synthesis of the Cutting-Edge Research Literature. *Journal of Computer and Education Research, 12*(24), 600-637.

Morales-Garcia, L., Navarro, C., & García, M.D.S. (2022). Epistemic and Mediational Suitability of Tasks Designed in a Mobile Learning Context. Eurasia *Journal of Mathematics, Science and Technology Education, 18*(3). Scopus. https://doi.org/10.29333/EJMSTE/11708

Russell, S. J., & Norvig, P. (2020). Artificial intelligence: A modern approach (4th ed.). Pearson.

Tunjera, N., & Chigona, A. (2020). Teacher Educators' Appropriation of TPACK-SAMR Models for 21st Century Pre-Service Teacher Preparation. *International Journal of Information and Communication Technology Education (IJICTE), 16*(3), 126–140. https://doi.org/10.4018/IJICTE.2020070110

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008. https://arxiv.org/abs/1706.03762

*Younis, B. (2024). Effectiveness of a professional development program based on the instructional design framework for AI literacy in developing AI literacy skills among pre-service teachers. *Journal of Digital Learning in Teacher Education*, *40*(3), 142-158.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0